An Operational Assessment of Collaborative Decision Making in Air Traffic Management:

"Measuring User Impacts through Performance Metrics"

Final Report

February 2000

Document Control Number: R90145-05

Project Lead

The Federal Aviation Administration

David Knorr James Wetherly

Project Team

University of Maryland

Michael Ball
Robert Hoffman
Tasha Inniss

Metron Inc.

Midori Tanino
Lara Shisler
Michael Brennan

MITRE/CAASD

Michelle Blucher
James DeArmon
Amy Miyoko Darrow

John A. Volpe Center

Ken Howard
Eugene Gilbo

GEMS Inc.

Richard Danz

TRW/SETA

Jeffrey Biros

Acknowledgements

This document is a collaborative effort incorporating the expertise of many organizations in the Federal Aviation Administration (FAA) and aviation industry. The FAA organizations included the Free Flight Phase 1 Program Office (AOZ), The Investment Analysis and Operations Research Branch (ASD), and the Air Traffic Control System Command Center (ATCSCC). Industry organizations included the Air Transport Association (ATA), the National Center of Excellence for Aviation Operations Research (NEXTOR), the John A. Volpe National Transportation Systems Center, Metron Inc., MITRE's Center for Advanced Aviation System Development (CAASD), SETA (Systems Engineering and Technical Assistance), and CALIBRE Systems.

Specifically, we would like to thank the various contributors to this project that attended the project meetings, shared their knowledge and expertise, and provided feedback to the preliminary analyses. These include Forrest Terral, Joseph Rather, and Anthony Milligan (ATCSCC), Miro Lehke (ATA), Daniel Citrenbaum, Charles Thorpe (FAA/ASD), Mike Wambsganss, and Terry Thompson (Metron, Inc.), Richard Oiesen (Volpe), Robert Juliano (SETA), and Bruce MacDonald (CALIBRE Systems).

We would like to also recognize and thank several U.S. airlines that assisted in this assessment by providing feedback to the surveys that were distributed. These include Transworld Airlines, Business Express Airlines, Midwest Express Airlines, and Federal Express.

We apologize if we missed any other persons or organizations that may have contributed to this report.

Executive Summary

Background

Free Flight Phase 1 (FFP1) is a collaborative effort between the FAA and industry to provide early benefits to users of the National Airspace System (NAS) through the deployment of core capabilities at selected sites. Capabilities and sites for FFP1 were agreed to by the FAA and industry (represented by RTCA) as documented in Addendum 1 of the Government/Industry Operational Concept for the Evolution of Free Flight, prepared by the RTCA.[1]

Collaborative Decision Making (CDM) is one of five core capabilities incorporated into FFP1. Each of these five capabilities[2] has operated in prototype form and is considered proven technology. The challenge for FFP1 is to replicate these operational capabilities at each of the selected sites and evaluate their effectiveness by the year 2002. To facilitate this challenge, the FFP1 office and the RTCA collaborated on an operational impact evaluation plan for each capability.

CDM is the most mature of FFP1's core capabilities. As of September 1999, the major capability of CDM, Ground Delay Program Enhancements (GDP-E), has been operating at all major airports in the (contiguous) United States for more than one year. (CDM's other capabilities, Collaborative Routing [CR], and NAS Status Information [NASSI] are nearing completion of their FFP1 goals.) Because of the existence of sufficient operational data, we were able to perform a rigorous evaluation of CDM's operational impact focusing primarily on GDP-E.

Many anecdotes exist that describe situations where GDP-E has contributed to millions of dollars in airline savings. GDP-E has also contributed to improvements in data quality, timeliness, and slot allocation under ground delay program conditions. As a result, CDM has become widely supported by both the FAA and the user community. This report documents the quantitative analyses used to measure CDM's operational impact on the National Airspace System (NAS) and proposes measures/metrics for continued evaluation of CDM's progress.

The analytical approach established in collaboration with the RTCA is to compare representative metrics of operational performance both pre- and post- implementation. The intent is to identify quantitative evidence that desired changes in operational performance have been achieved. Also provided in this report are the analyses of several metrics that provide general measures for analyzing continuous improvements in tool performance following deployment.

With the highly dynamic environment of the NAS, the Metrics Team recognized the challenges inherent in a pre- versus post- comparison. To address this type of challenge,

---

[1] Addendum 1: FFP1 Limited Deployment of Select Capabilities is a report of the Government/Industry Operational Concept for the Evolution of Free Flight and was prepared by the RTCA Select Committee.
[2] Free Flight Phase 1 capabilities include Surface Movement Advisor (SMA), Traffic Management Advisor (TMA), passive Final Approach Spacing Tool (pFAST), User Request Evaluation Tool (URET), and Collaborative Decision Making (CDM).

analysis methodologies can become quite complex. For this reason, estimates of the value of the capability without specific comparisons with pre- and post- data are employed to provide an uncomplicated measure of GDP-E operational value. For example, compression benefits have been frequently referenced as estimates of GDP-E value and have become a widely accepted uniform measurement of system performance.

Introduction to Collaborative Decision Making

Collaborative Decision Making (CDM) was conceived out of the FAA's Airline Data Exchange (FADE) experiments that began in 1993. These experiments proved that having airlines submit real-time operational information to the FAA could improve air traffic management decision-making. CDM is an effort to improve air traffic management through information exchange, procedural improvements, tool development, and common situational awareness.

The initial focus of CDM, known as Ground Delay Program Enhancements (GDP-E), began its prototype operations at San Francisco (SFO) and Newark (EWR) airports in January of 1998. Under GDP-E, participating airlines send operational schedules and changes to schedules to the Air Traffic Control Systems Command Center (ATCSCC) on a continual basis. This schedule information includes, but is not limited to, flight delay information, cancellations, and newly created flights. Through the use of the Flight Schedule Monitor (FSM), the ATCSCC uses this information to monitor airport arrival demand and to conduct ground delay programs (GDPs). The airlines are also able to monitor arrival demands and model ground delay programs via FSM but do not have the capability to alter or implement ground delay programs.

In addition to improving the execution of GDPs, CDM has been found to have application to other air traffic management problems, such as airspace congestion due to heavy traffic or En-Route weather. CDM's Collaborative Routing function is intended to provide better information to airspace users about potential flow problems that are likely to require rerouting or other flow management actions. This may allow users to prepare for possible effects on their operation in advance. The National Air Space Status Information function will provide a mechanism to share critical safety and efficiency data with NAS users.

Summary of CDM Performance Metrics and Their Valuations

Improved Data Quality

CDM has produced new information by combining FAA and airline data sources. All CDM airline participants have implemented data feeds from their operations systems into the CDMnet. Using these data feeds, the airlines provide information on flight cancellations, mechanical delays, and other events that impact the demand on the NAS. This information is merged with FAA-generated information by systems at the Volpe Center into a real-time data feed, known as the "CDM String."

Through the CDMnet, the CDM-enhanced information has been distributed in an unprecedented fashion. In fact, probably the most significant aspect of the new CDM information infrastructure is that the Airline Operations Centers (AOCs) receive the same information and decision support tools as do FAA ATCSCC specialists. Such information is critical in enabling airline operations specialists to plan responses to changing conditions and possible FAA control actions. Previously, such information was not available to airline operations planners or was only available "after-the-fact," when it could no longer be used to influence decision-making.

Our analyses have found that the information flowing over the CDM string is of higher quality - greatly improving NAS system predictability. Moreover, we have found that the improvements are most dramatic when conditions induce irregular system operations. The cancellation notice and integrated predictive error (IPE) metrics identified in NEXTOR's August 1998 report provide evidence that CDM has made improvements to data quality, specifically with respect to departure predictions and flight cancellations.

Improved EDCT Compliance

The terms EDCT (Estimated Departure Clearance Time) and CTD (Controlled Time of Departure) have the same meaning: the FAA-assigned time at which a flight is supposed to depart under a ground delay program. The successful execution of a GDP depends heavily on departure compliance. However, failure to comply with the EDCT during GDPs has been a problem for years. CDM has been providing airlines with real-time airport arrival information and has encouraged airlines to focus on EDCT compliance in a collaborative manner.

We have found that departure compliance has improved significantly under CDM. The average on-time departure percentage has increased from 50.85 to 65.87 percent of all flights issued an EDCT. This means that 15.02 percent more flights maintain departure compliance since the inception of CDM. This implies that the number of on-time departures has grown by 29.54 percent.[3]

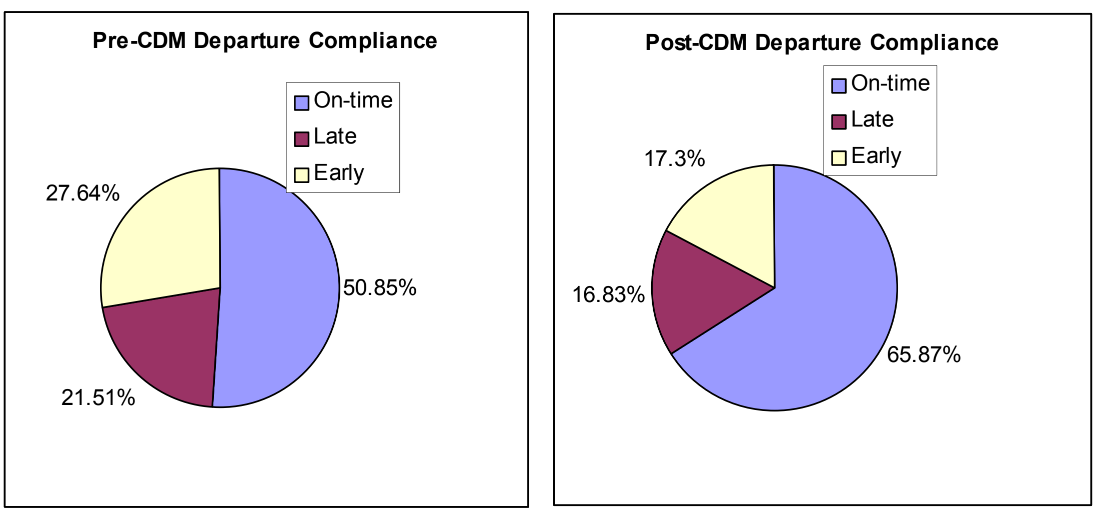

Figure I. Pre- and Post- CDM Depature Compliance

[3] Taken with respect to the original percentage of on-time flights (i.e., 29.54 = 100 × [15.02 / 50.85]).

The 15.02 percent growth in on-time flights can be decomposed into a 10.34 percent reduction in early departures and a 4.68 percent reduction in late departures (15.02 = 10.34 + 4.68). These correspond to 21.76 and 37.41 percent reductions in the respective categories.[4] Airlines have always had an incentive to reduce late departures regardless of CDM status. This may be the reason why we see less improvement in this category compared to improvement in early departures. Nevertheless, the 21.76 percent late departure improvement over the pre-CDM period is a significant achievement. The improvement in early departures over the pre-CDM period is 37.41 percent. We believe that this improvement is resulting from an active information exchange between the FAA and airlines and improved attention toward flight operations.

Improved Predictability: Integrated Predictive Error - The IPE Metric

CDM has made a concerted effort to improve the accuracy of flight departure predictions. Participating air carriers have voluntarily augmented ETMS flight data with their own departure predictions. The premise is that each airline has the most complete picture of its operations (delays due to connectivity, gates, etc.), thus enabling it to make more accurate predictions of its departure times than ETMS.

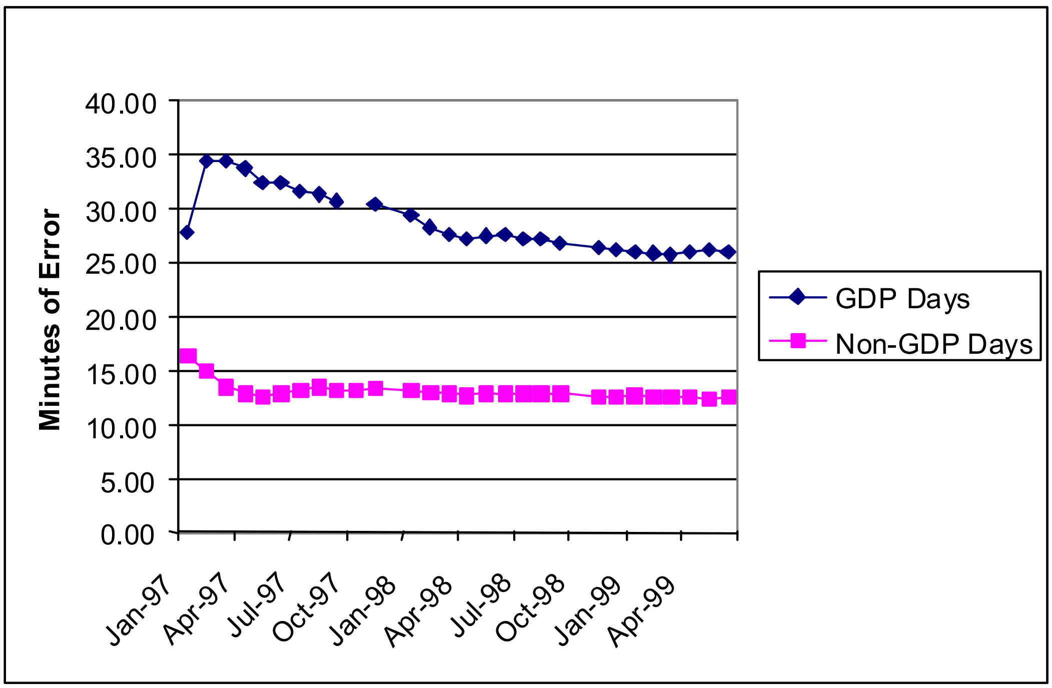

Figure II. IPE-6 Metric Trends for Departure Predictions (Cumulative Average Since Jan –97)

[4] The percentage of early flights decreased from 27.64% of all flights pre-CDM to 17.30% of all flights post-CDM, for a change of 27.64 – 17.30 = 10.34 percent. With respect to the pre-CDM statistic of 27.64%, this implies a change of 100*(10.34/27.64) = 37.41 percent. Similarly, pre-CDM and post-CDM percentages (over all flights) of late departures were 21.51 and 16.83 for a drop of 21.51 – 16.83 = 4.68 percent of all flights and a change of 100*(4.68/21.51) = 21.76 percent.

We used the integrated predictive error (IPE) metric to monitor long-term trends in flight departure predictive accuracy. IPE is a weighted average of the errors in a stream of predictions made over time for a single event. Based on the data used in this analysis we have found that, on average, departure prediction accuracy increases (has less error) as a departure flight approaches. Since August of 1997, the average departure prediction error on GDP days at San Francisco Airport (SFO) has dropped from 31.29 minutes per flight to 26.06 minutes per flight, for a net reduction of 5.23 minutes per flight.[5] Comparable results have been found at Newark Airport (EWR).

The dramatic improvement on GDP days is noteworthy because accurate flight data is most crucial during a GDP. Average IPE values for non-GDP days at SFO and EWR have dropped as well. In fact, for both airports, the departure prediction error has been pushed below 15 minutes, the industry-wide standard for an on-time event.

Enhanced GDP Performance: The Rate Control Index (RCI)

The rate control index (RCI) measures the flow of air traffic into an airport and compares it to the targeted flow that was set by the traffic flow managers at the ATCSCC during a ground delay program. In other words, it is a measure of how well we executed the planning for a GDP. A single index, or percentage, is reported for the entire performance of a GDP on a single day. A higher score (e.g., 95 percent) corresponds to better performance, meaning the flow of traffic into the airport closely matched the targeted flow of traffic, both in quantity and in distribution. The RCI metric is notable because it is designed to assess the *execution* of a program rather than the retrospective appropriateness of the plan underlying the program. When applied to traffic flow at the terminal space prior to any airborne holding, the RCI metric is virtually independent of the program goals set or the accuracy of the weather forecasts upon which it is based. In particular, RCI is adept at flagging GDPs with particularly high or low performance.

We tracked results of the RCI metric over a 30-month period for traffic flow into the terminal space of San Francisco (SFO) and Newark (EWR) airports, these being the two original CDM prototype operations airports. We found that traffic flow into both airports had improved slightly, more so at EWR than at SFO, meaning that the rate of flow tends to match more closely the targeted flow than it has in the past. In general, there tends to be more variation at EWR than at SFO. We attribute this to the complexity of EWR's terminal space (bordering on different traffic centers) and the less predictable nature of East Coast traffic. Also, we caution that the results at EWR are less conclusive than at SFO because the computation of this metric is dependent upon the modeling of airborne holding, which is more difficult at EWR than at SFO.

The integrity of some of the data (mainly AZ messages) underlying the RCI metric and the airborne holding models upon which it relies are being further examined. At a later time, when more data has been collected for other airports, further conclusions can be reached about traffic flow performance during a GDP.

---

[5] Taken as a cumulative average of all months since January of 1997. Departure predictions were tracked over a six-hour period prior to departure. See the body of the report for monthly statistics and statistical methods.

Reduced Near-Term GDP Cancellations

A *near-term* cancellation of a GDP is when a GDP is aborted within 30 minutes of its planned start time (the time at which the first controlled flight is scheduled to land). Since ground delay impacts flights prior to their airport departure, many flights will have absorbed delays well in advance of the start time of the GDP. Thus, all assigned ground delays absorbed prior to the start of the canceled GDP are (in hindsight) unnecessary. For this reason, near-term cancellations of GDPs are considered undesirable.

We tracked the number of instances of near-term GDP cancellations both pre- and post-CDM at six major airports. We conjectured that the combination of improved demand information and the power run feature of FSM that allows ATCSCC personnel to delay the implementation of a GDP to the last possible minute should decrease the number of near-term cancellations. Some airports showed improvement - others did not. Most notably, there has been a remarkable improvement at St. Louis in the percentage of near-term GDP cancellations. We believe that this is the result of superior data quality of the two major airlines that dominate the airport. This caliber of data quality is, in turn, attributed to the use of *daily download*, the replacement of (often) obsolete OAG information with fresh airline operational data at the start of each day (not all carriers participate in daily download). In addition, these airlines have provided positive feedback on FSM and the procedures adopted for CDM.

User Anecdotes on GDP-E Benefits

One of the methods we used for assessing the benefits of CDM is collecting anecdotal evidence from both the user community and the ATCSCC. This uncovers positive experiences of CDM participants and benefits of CDM that may be missed by quantitative analysis. The method was tailored to each of the user groups. For the air carriers, a 10-question survey was distributed by e-mail. For the ATCSCC, anecdotal evidence was collected from the ATCSCC specialists' logs, mostly in the GDP critiques. Subsequent responses and comments from the ATCSCC logs are included in Appendix C.

Compression Benefits

Compression is an inter-airline resource allocation algorithm that advances take-off times of flights to fill arrival slots vacated by cancelled or delayed flights. This makes more efficient use of airport arrival resources by utilizing arrival slots that would have been unused through initial slot allocation or the intra-airline substitution process. This reduces the number of minutes of planned (FAA-assigned) ground delay. Compression, which was introduced by the GDP enhancements of CDM, has proven to provide substantial benefits to the user community. Between September 8, 1998 and August 31, 1999, there have been a total of 3,735,427 cumulative minutes of assigned ground delay reduced due to compression (and over 4 million cumulative minutes since the start of prototype operations). The benefits of these savings go beyond just averting needless ground delay. Compression provides the ATCSCC with a tool which helps create a smooth arrival rate into an airport, without wasting valuable arrival resources. As a result, the ATCSCC has more timely and accurate information about cancellations and delays. This allows the airlines and the ATCSCC to compress open slots (resulting from cancellations) that are not utilized through the substitution process.

Increased User Equity

Enhancements to GDPs introduced a new process for making the initial assignment of flights to arrival slots during a ground delay program. Through experimentation and dialogue, the air carriers and the FAA have worked hard to make this rationing process equitable to all parties involved in a ground delay program. The result of their efforts is an algorithm called Ration-by-Schedule (RBS). RBS rations arrival slots according to scheduled arrival times as posted in the OAG, as opposed to real-time, estimated arrival times. This removes disincentives for airlines to notify the ATCSCC of delays and establishes the concept of slot ownership. We designed four metrics to assess the equity of the current arrival slot allocation process. Based on an evaluation of these metrics, the RBS algorithm has proven to be a fair and equitable mechanism for assigning arrival slots to flights during a GDP. Moreover, the decision support tools embedded within FSM provide GDP equity statistics which may be used by ATCSCC specialists' to model various GDP options.

Tailored GDP's through Revisions

The modification of GDP parameters such as scope, duration, or the associated airport acceptance rate (AAR) is known as a *revision*. Prior to CDM, the ATCSCC did not have the capability to revise a program once it was in effect. While they did have the ability to affect GDP-controlled traffic flow by means such as blanket delays (adding a fixed number of minutes of delay to all flights), the methods for program modification were cruder and less effective than the revision capability now provided by CDM. According to Forrest Terral of the ATCSCC, changes to programs prior to CDM have, in many cases, resulted in the loss of program integrity.

One of the most powerful revision that can be made to a GDP is to extend the length of a program. This allows the ATCSCC to control later-arriving traffic when adverse weather effects last longer than expected, and to smooth out pent-up demand (a *stack*) that may accumulate toward the end of a program.

Since GDP revision was not an option prior to CDM, we were not able to make a pre- vs. post-CDM analysis of the effect of this tool. However, we can state that this tool has been used frequently since the inception of CDM and has proven to be highly effective for controlling traffic flow. At least 10 log entries in the ATCSCC GDP critique attest to the effectiveness of revisions to smooth out the traffic (and reduce departure delays.) The flexibility of this tool has resulted in the fuller use of capacity and a reduction in airborne holding.

Conclusions

CDM GDPE has been operating for more than one year (in the contiguous U.S.). It has become part of the ATC management system and has been widely supported by both the FAA and the user community. We can reason that the ATC system operates "better" with this improved data quality, timeliness, and slot allocation processes. Like any system, GDP-E can be improved further. Measures like IPE, RCI, and compression with associated equity can help us better understand the value of these improvements.

This analytic team has sought a unified approach for analyzing CDM performance. The most natural candidate, and the one that we have tried to pursue, is to collect performance statistics both before and after program implementation and review them for positive trends. However, this type of pre- versus post- analysis assumes that all else remains constant during the implementation or that, at the very least, results can be screened for complicating factors. But the NAS is a highly dynamic environment that does not lend itself to controlled experiments. For several of the metrics, data was not available in the pre-implementation phase. As a result, this pre- versus post- analysis was pursued only when it made sense to do so.

Initially, we had hoped to adopt simplistic metrics to answer 'bottom-line' questions such as: "has throughput improved?," "have travel times been reduced?," and "are there fewer delays?" Many of these metrics failed for a variety of technical reasons including incomplete, corrupted, or altogether unavailable data. For other metrics, we found that the further we probed, the less certain we were of the impact that CDM should have. For instance, should the number of ground delay programs increase or decrease as a result of CDM? On the one hand, airline-supplied cancellation data should provide more accurate arrival demand predictions and therefore avoid superfluous ground control actions. (This has been documented on several occasions.) On the other hand, the refinement of GDP procedures and the ability to revise parameter settings has provided the community with a more equitable, more resilient tool. Wider acceptance may have encouraged more frequent use. (Indeed, there have been cases in which the air carriers have requested the use of a GDP.)

It is far easier to assess the programs within a corporate environment. Value can be measured simply by the difference of dollars invested and dollars saved or generated. Unfortunately, we don't have the same unifying objective (profit) or standard of measurement (dollars) in an aviation setting. True, some assessment can be made of air carrier dollars saved in the form of fuel, man-hours, etc. But fuel consumption is not always a primary concern; an airline may knowingly trade fuel consumption for an earlier arrival or a passenger comfort. We have learned that these types of computations are highly conditional - both situation specific and air carrier dependent.

Many FAA systems have sought to ensure safety *and* improve efficiency within the NAS. Since air traffic safety within the United States has an impressive track record, this is not one of the areas that CDM has sought to directly improve. As far as efficiency is concerned, this is generally taken to mean efficient use of NAS resources on a *global* scale, without reference to individual organizations using those resources. But what's inefficient for one user may profit another. The NAS serves the needs of many groups whose objectives change on a daily basis. This makes it hard to find a uniformly applicable metric. In fact, the very premise of measuring economic savings on behalf of the air carriers defies one of the cornerstones of CDM: that each airline should weigh for itself the economic value of alternative actions.

One uniform measurement of progress that the CDM community has agreed upon is minutes of (FAA) planned ground delay that has been avoided as a result of the Compression algorithm housed in ground delay program procedures. This has become the showcase achievement of CDM activities. As of this writing, the CDM program has logged over 4,000,000 minutes in savings of planned ground delay. Not all of this may transfer to minutes saved in gate-to-gate time; ten minutes of savings at the origin airport could be offset by ten minutes in airborne holding at the destination airport. Nonetheless, the opinion of CDM experts is that the true savings from compression is a significant fraction (i.e., about one-half) of the 4,000,000 minutes.[6]

Weather has proven to be the primary culprit in the disruption of airline operations and the foremost cause of FAA-induced control actions. Ironically, this has also proven to be the prime obstacle for analysis of aviation operations. The RCI that we have developed is the first metric to rigorously factor out weather conditions from evaluation of large-scale, multi-participant performance.

Second only to weather, the greatest impediment to this analysis was the difficulty of modeling what *would have* happened had CDM not been in place. The modeling of something as seemingly simple as the number of aircraft that could have landed under different circumstances requires not only highly sophisticated modeling techniques but also the input of dozens (or even hundreds) of parameters, many of which have to be subjectively fixed. This has, in many cases, caused us to significantly alter or even abandon our first-choice metrics. We have guarded against modifying or distorting metrics to the point that the results become uninterpretable. Whenever possible, we have formulated the simplest, most appropriate metric, given the current field of knowledge, and we have documented the reasons that other metrics were deemed inadequate or inappropriate.

The primary mission of this team has been the development and application of CDM-related metrics. But it is incumbent upon us to place our findings in perspective with CDM activities and accomplishments. What has started out as a grass roots effort has developed into the most mature FFP1 core capability. The true benefits of CDM may lie in the communication and awareness that has been established, particularly between the users of the NAS and its service providers. The working relationships and cooperative efforts that it has formed are essential for the promotion of free flight.

Its strength lies primarily in its ability to surmount cultural behaviors that inhibit communication between entities. CDM has managed to consistently deliver on its short-term objectives while maintaining the pursuit for long-term solutions. Specifically, CDM has:

- Created much needed data sources through the submission of real-time airline operational data;
- Established a communications infrastructure, the CDMnet, for the promotion and dissemination of situational awareness amongst all parties in the NAS;

[6] Based on the commonly accepted (perhaps dated) industry standard of $25.00 per minute, this is equivalent to $50,000,000 in savings to the airline community since January of 1998.

- Provided equity and predictability to ground delay program procedures;
- Provided an open forum for problem identification and problem solving through the participation and coordination of airlines, industry, and the FAA.

To whatever extent possible, we have tailored the body of CDM metrics to reflect CDM activities and accomplishments. The body of metrics continues to grow and evolve with the CDM program. The technological solutions for improving the NAS have in many cases outgrown our ability to assess their impact. This report is as much a treatise on aviation performance metrics as it is an assessment of CDM program status. But much headway can be made in the development of metrics and analytic methods. For continued analysis of CDM program performance the following metrics should be considered:

- Rate Control Index (RCI)
- Integrated Predictive Error (IPE)
- Compression
- Equity
- Slot usage

Although not investigated in this document, usage of CDM decision support tools such as Flight Schedule Monitor (FSM) can further assist decision-making and operations analysis.

It is vital that aviation performance metrics and solutions to problems within the NAS be further developed in a complementary fashion. These will be applicable not only to the CDM program, but other FFP1 programs as well.

Table of Contents

Section **Page**

List of Figures

List of Tables

Table **Page**

1.0 Introduction

Free Flight Phase 1 (FFP1) is a collaborative effort between the FAA and Industry to provide early user benefits to users of the National Airspace System (NAS) through the deployment of five core capabilities at a limited number of sites.

Prior to the creation of FFP1 each of the five core capabilities[7] was operating in prototype form. Of these capabilities, Collaborative Decision Making's (CDM's) ground delay program enhancements (GDP-E) is the most mature of all FFP1 systems and as of October 1999 it was operating nationally for more than one year. For this reason, sufficient data exists to evaluate GDP-E impact on NAS operations.

As part of FFP1, each capability is to be evaluated for a minimum of one year after deployment in order to assess the benefit of a national deployment. The analytical approach established with the RTCA was to compare representative metrics of desired impacts both pre- and post-implementation. The intent was to identify quantitative evidence that desired changes in operational performance have been achieved.

During the evaluation effort, it became obvious that for some of the original RTCA metrics, insufficient pre- GDP-E data was available for several of the metrics to make convincing pre versus post comparisons. Additionally, the analysis team recognized the complexity in interpreting several performance metrics where the preferred direction of change is unclear (i.e., is an increase in aircraft substitutions during GDPs "good" or "bad"). The analysis methodologies of these metrics are documented in Appendix A.

Historically, compression savings has represented GDP-E performance improvements because of its direct relationship to delay savings and capacity increases. In this report several other metric and methodologies are identified to better evaluate and understand GDP-E functionality and performance impacts. This report documents these measures and how they may be used to continuously measure trends in GDP-E operational performance as the capability matures.

1.1 Background and Structure of Evaluation

In August 1998, The National Center of Excellence for Aviation Operations Research (NEXTOR) released a benefits assessment on CDM. This preliminary report highlights several areas that have produced tangible benefits as well as areas that are prospects for future benefits. It later proved to be a valuable resource to the development of a thorough GDP-E evaluation plan. As a result, several of the metrics noted in this evaluation plan have been either derived or directed from the NEXTOR report.

In February 1999, a subgroup of the CDM community including the Free Flight Phase 1 Program Office (AOZ), Volpe, ASD-400, NEXTOR, Metron, Inc., TRW/SETA, and MITRE/CAASD began identifying metrics and developing methodologies to determine operational impacts of GDP-E and its components on the National Airspace System (NAS). (Since the fundamental principal underlying GDP-E [and CDM] is the

[7] Free Flight Phase 1 capabilities include Surface Movement Advisor (SMA), Traffic Management Advisor (TMA), passive Final Approach Spacing Tool (pFAST), User Request Evaluation Tool (URET), and Collaborative Decision Making (CDM).

improvement of information and the sharing of data, it is important for the CDM section of the evaluation plan to incorporate the knowledge, ideas, and experience of all groups that are functionally involved.) Since February, the subgroup has met several times, approximately every two months, to accomplish this task.

This effort began with the review of the CDM performance metrics identified by the RTCA Working Group. These metrics were then combined to a number of metrics identified by the CDM subgroup. This consolidated list of CDM performance metrics was the starting point with which the subgroup began their analysis. Although each metric was considered, the subgroup was uncertain about the value of several of these metrics for effectively measuring impact to the NAS. Other metrics such as NEXTOR's integrated predictive error (IPE) and rate control index (RCI) were assumed sound since analysis was well underway.

During the next several months, members of the subgroup performed preliminary analysis of their assigned performance metrics. Once completed, the analysts were able to determine if their assigned metric was viable or impractical. From these evaluations, several metrics were pulled from the consolidated list. Several of the original metrics were not thoroughly analyzed for reasons other than the inability of the metric to potentially measure NAS performance. In fact, these metrics would likely provide a sure measure of CDM performance however, due to limited resources or complexity of analysis, they were not considered in this report. A detailed explanation of why these metrics were not the best indicators of GDP-E performance is included in Appendix A.

On June 7, 1999, the CDM subgroup agreed to collaborate their efforts on a written report incorporating analysis of the various CDM performance metrics as well as providing the current status of the various CDM elements including Initial Collaborative Routing (ICR) and National Airspace Status Information (NASSI). This report focuses on metrics that were representative of the RTCA's original intent to quantify operational impacts of GDP-E.

1.2 Collaborative Decision Making

Collaborative Decision Making (CDM) was conceived out of the FAA's Airline Data Exchange (FADE) experiments that began in 1993. These experiments proved that having airlines send updated schedule information to the FAA could improve air traffic management decision-making. CDM has evolved from these same principles in an effort to improve air traffic management through information exchange and data sharing.

Ground Delay Program Enhancements (GDP-E), the initial focus of CDM, started prototype operations at San Francisco (SFO) and Newark (EWR) airports in January 1998. Under GDP-E, participating airlines send operational schedules and changes to schedules to the Air Traffic Control Systems Command Center (ATCSCC) on a continuous basis. This schedule information includes, but is not limited to, flight delay information, cancellations, and newly created flights. The ATCSCC uses this information to better implement and manage ground delay programs (GDPs).

In addition to improving the execution of GDPs, CDM has been found to have application to other air traffic management problems, such as airspace congestion due to heavy traffic or En-Route weather. CDM's Collaborative Routing (CR) function is intended to provide better information to airspace users about potential flow problems that are likely to require rerouting or other flow management actions. This may allow users to prepare for possible effects on their operation in advance. The National Air Space Status Information (NASSI) function will provide a mechanism to share critical safety and efficiency data with NAS users. A working group has been formed that will determine what these data are and how to set priorities for distributing the data.

1.3 Current Status of CDM

The initial operational implementation of CDM has been aimed at the development of new operational procedures and decision support tools for implementing and managing ground delay programs (GDPs). However, a much broader class of problems in air traffic management is envisioned for the application of CDM. The current status of CDM can be summarized as follows:

Agreement on New GDP Paradigm

The airlines and the FAA have agreed on a new resource allocation concept embodied within two procedures: ration-by-schedule and compression. These provide mechanisms for a "fair" rationing of arrival time slots under degraded conditions. A key driver to all of CDM is that all parties have agreed on the fairness of the allocation. Moreover, the ability of a specific airline to achieve its fair allocation is not hindered by that airline providing the most up-to-date information on the status of its flights.

Regular Meetings of all CDM Players

All CDM players, including airline representatives, FAA traffic flow managers, developers and planners, software developers, consultants, and researchers, meet about every two months which has greatly assisted the productivity of team's efforts. These meetings have allowed for the CDM agenda to be pushed forward very rapidly and have led to greater mutual understanding among the parties involved. Several subgroups pursue specialized agendas. These smaller groups typically hold independent meetings.

Flight Schedule Monitor (FSM): The CDM Decision Support Tool

FSM has undergone an evolutionary development over the past few years and it now embodies the requirements expressed by both FAA and airline users. The tool is in place at both the ATCSCC and the operations centers of all CDM participating airlines. Plans are in place to distribute FSM to other FAA facilities.

CDMnet

The participating airline operational control centers (AOCs), the ATCSCC, the Volpe National Transportation Systems Center (the hubsite of the Enhanced Traffic Management System [ETMS]). Metron (the developer of FSM) as well as certain other parties are all interconnected via a private network, the CDMnet. This network is used to exchange CDM operational information.

GDP Implementation

The initial implementation of CDM commenced in January of 1998. All participating airlines, the Volpe Center and the ATCSCC have been exchanging information via the CDMnet. Furthermore, starting on January 20, 1998, CDM procedures were used to implement all GDPs at San Francisco and Newark airports. On April 20, 1998, this list was expanded to include La Guardia and St. Louis airports. On September 8, 1998, use of CDM GDP procedures was expanded to all US airports.

Initial Collaborative Routing

Early on it was recognized that CDM had applicability to enroute airspace management and a subgroup has been actively pursuing this area. A set of Initial Collaborative Routing (ICR) tools and procedures were prototyped and tested during the 1999 severe weather season. These included, national CRCT (Collaborative Routing Coordination Tool), a tool developed by the MITRE Corporation, which supports the identification of airspace congestion and the development and evaluation of alternative routes; CCFP (Collaborative Convective Forecast Product), a national convective weather forecast, which represents a consensus based on inputs from AOC and ARTCC weather units; use of information and application distribution products (PictureTel and World Wide Web) to support CR decision making; LAADR (Low Altitude Arrival and Departure Routes), which embodies a set of procedures for allowing the use of low altitude routes in order to avoid congested airspace; CDR Coded Departure Routes, which involves a set of procedures and a database for the creation, storage and dissemination of alternate routes to be used in order to avoid airspace blocked by severe weather.

NAS Status Information Dissemination

A CDM National Airspace System Status Information (NASSI) subgroup has been established to identify that information on NAS conditions that would benefit the FAA and air carriers for strategic planning and general awareness purposes and to make recommendations for its dissemination in real or near real time. For airport data, six categories have been identified: GDP Projected Demand and Capacity, Departure Delays, Planned & Actual Pushback Times, Airport Acceptance Rates, and Airport Configurations. For airspace data, four categories have been defined: SUA via SAMS, Miles-in-Trail (MIT) Restrictions, Arrival Delays, Severe Weather Avoidance Program (SWAP) Routes, and Runway Visual Range. Data from virtually all of these ten categories are now available to the FAA and industry via the CDMnet.

2.0 Performance Metric Indicators for Operational Impact

2.1 Data Quality

The use of CDM has produced new information by combining FAA and airline data sources. All CDM airline participants, including American Airlines, Continental Airlines, Delta Airlines, Northwest Airlines, Southwest Airlines, Trans World Airlines, United Airlines and US Airways have implemented data feeds from their operations systems into the AOCnet. Using these data feeds, the airlines provide information on flight cancellations, mechanical delays, and other events that impact the demand on the NAS. This information is merged with FAA generated information by systems at the Volpe Center into a real-time data feed, known as the "CDM String."

Through the CDMnet, the CDM-enhanced information has been distributed in an unprecedented fashion. In fact, probably the most significant aspect of the new CDM information infrastructure is that the airline operations centers receive the same information as the FAA ATCSCC specialists. Such information is critical in enabling airline operations specialists to plan responses to changing conditions and possible FAA control actions. Previously, such information was not available to airline operations planners or was only available "after-the-fact," when it could no longer be used to influence decision-making.

Our analyses have found that the information flowing over the CDM string is of higher quality. Moreover, we have found that the improvements are most dramatic when the system is under stress and the information is most critically needed. The performance metrics integrated predictive error (IPE) and rate control index (RCI) noted in this report, and identified in NEXTOR's August 1998 report, provide evidence that suggests improvements in data quality, specifically improvements in departure prediction and cancellations.

2.2 EDCT Compliance

The terms, EDCT (Estimated Departure Clearance Time) and CTD (Controlled Time of Departure), have the same meaning. Both EDCT and CTD times are the departure times under a GDP. It is highly important for GDP-affected flights to comply with these times since the successful execution of GDPs depends heavily on departure compliance. However, failure to comply with the EDCT during GDPs has been a known problem for years. CDM has been providing airlines with real-time airport arrival information and has encouraged airlines to focus on EDCT compliance in a collaborative manner. This study presents the effect of GDPE on the air traffic system from an EDCT compliance perspective. We believe that if CDM is working, EDCT compliance should be improving.

The period considered for the EDCT compliance study is from February 1996 to July 1999. The baseline period is February 1996 through December 1997. The period between January 1998 and August 1998 is the transition period. Specifically, the CDM GDP experiment began on January 20, 1998 with two airports, San Francisco and Newark. The number of CDM GDP airports was expanded to include La Guardia and St.

Louis, on April 28, 1998. The CDM period, prototype operations at all airports, is from September 8, 1998 to July 1999.

As for the data source, we used historical FSM data. The historical FSM data in 1996 was based on the ETMS test string data. The FSM historical data under went the transition period during 1997. In this year, Volpe Transportation Systems Center (TSC) created the CDM string, which is based on the ETMS string and includes airline data. We started to collect CDM data as Volpe's ADL project matured. Because of this, 1997 historical FSM data was based on mostly ETMS data and some ADL (CDM) data. Since 1998, FSM data has been purely based on CDM data. Because the EDCT compliance study required only two data fields, actual flight take off time (DZ) and EDCT (controlled departure time, CTD), the effect of the data source transition from ETMS to CDM is minimal.

We collected all flights that were issued an EDCT. Then we determined which flights had an enforced EDCT and which flights had a non-enforced EDCT. Some EDCTs are non-enforced due to the early GDP cancellation. For example, assume we had a GDP effective from 1400z to 1859z and that flight ABC123 had an EDCT of 1630z. The ATCSCC canceled this GDP at 1500z. At this point, ABC123 is no longer required to comply with 1630z EDCT. Such flights were removed from the study. We used only the flights affected by the active GDP for this study. If the GDP is prematurely canceled, we did not include the flights with EDCTs after the cancellation time.

We added one additional criterion to eliminate invalid data. If the difference between DZ and EDCT is outside of –120 and +180 minutes, the flight is either considered invalid or some special event took place for this particular flight and it is excluded from this analysis. This range was determined by reviewing historical data for inconsistencies.

The FAA definition of EDCT compliance is as follows:

1. Early Departure: the flight took off –6 minutes or earlier than its EDCT.
2. On-time Departure: the flight took off between -5 and +15 minutes of its EDCT.
3. Late Departure: the flight took off +16 minutes or later than its EDCT.

In some special cases, flights with an EDCT obtain a departure clearance time earlier than their EDCT over the telephone. Since we could only examine computer-collected data, it is possible that some flights that received departure clearance over the telephone are categorized as EDCT non-compliance. We believe the number of such occurrences is small and made our EDCT compliance results more conservative.

EDCT Compliance Results

Detailed EDCT Compliance data is displayed in Appendix E. It contains the month and year of the data followed by number and percentage of early departures, on-time departures and late departures.

Table 1 presents the information for *pre-CDM* and *CDM* GDP counts during 1998 when the transition between the two periods took place. We define the pre-CDM period as

February 1996 to June 1998 and August 1998. Technically, February 1996 to August 1998 should be the pre-CDM period. However, we did not include July 1999 because most GDPs (over 90 percent) in this month are CDM GDPs. The CDM period is the months between September 1998 and July 1999 and July 1998. CDM GDPs occur less than 90 percent of the time in the pre-CDM period; while in the CDM period, the percentage of CDM GDPs is greater than or equal to 90 percent. Although the CDM period started September 1998, we feel comfortable including July 1998 in the CDM period since more than 90 percent of the GDPs are CDM GDPs.

Table 1. GDP Counts and CDM GDP Percentages During the Transition Year

Date	Total No. of GDPs	Pre-CDM GDPs	CDM GDPs	CDM GDP %	Note
Jan-98	83	10	73	12%	CDM GDP started on 01/20/98 with EWR and SFO.
Feb-98	66	28	38	42%	
Mar-98	60	20	40	33%	
Apr-98	25	14	11	56%	LGA and STL are added to the CDM GDP airport list on 04/28/98.
May-98	41	26	15	63%	
Jun-98	40	25	15	63%	
Jul-98	21	19	2	90%	
Aug-98	28	18	10	64%	
Sep-98	33	32	1	97%	All contiguous US airports became GDM GDP airports on 09/08/98.
Oct-98	34	34	0	100%	
Nov-98	34	34	0	100%	
Dec-98	48	48	0	100%	

Table 1 presents the average percentages of early, on-time, and late departures for pre-CDM and CDM periods.

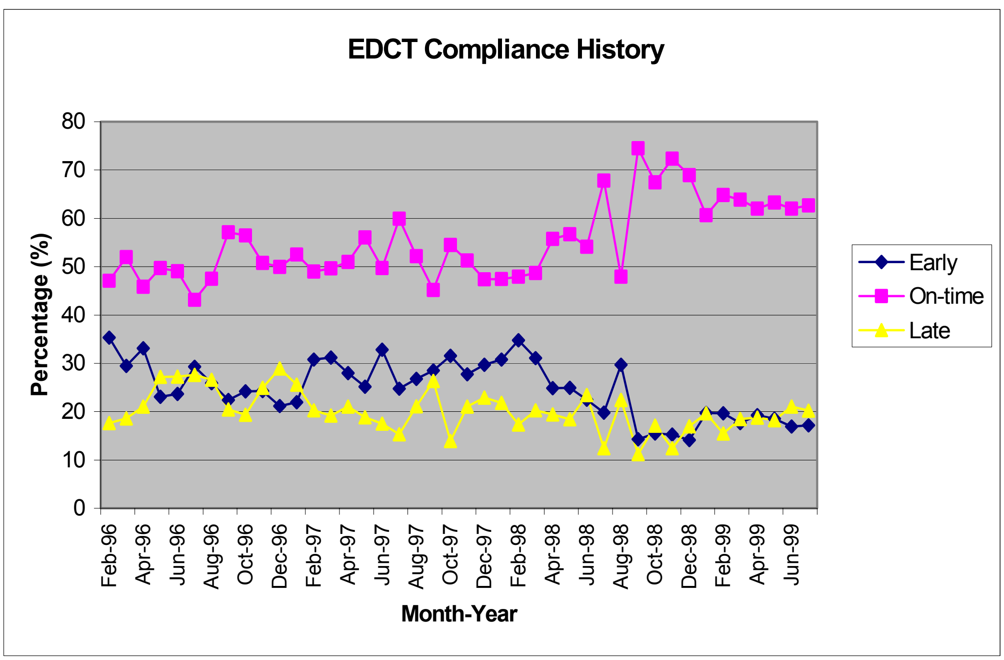

Figure 1. EDCT Compliance History

Figure 1 presents the EDCT Compliance history between February 1996 and July 1999. It indicates a steady improvement of EDCT compliance, starting in September 1998.

2.3 Predictive Accuracy of Flight Departures: The IPE Metric

CDM has made a concerted effort to improve the accuracy of flight departure predictions. Participating air carriers have voluntarily augmented ETMS flight data with their own departure predictions. The premise is that each airline has the most complete picture of its operations (delays due to connectivity, gates, etc.), thus enabling it to make more accurate predictions of its departure times than ETMS.

We used the integrated predictive error (IPE) metric to monitor long-term trends in flight departure predictive accuracy. IPE is a weighted average of the errors in a stream of predictions made over time for a single event. This number is called the IPE of the event. A numerical suffix indicates the number of hours over which the metric was tracked. IPE units are normalized by the tracking time and can be thought of as an average error (usually given in minutes). For instance, an ipe-6 value of 10 minutes would be obtained by making a steady stream of predictions over 6 hours each of which is off from the actual departure time by 10 minutes. See the Appendix in the 1998 NEXTOR report for a more detailed explanation of the IPE metric.

For each day between January 1997 and June 1999, we computed the average ipe-6 departure error over all flights bound for San Francisco (SFO) or Newark (EWR) airports. This process was repeated for the metrics ipe-5, ipe-4, ..., ipe-1, to arrive at six averages for each airport, for each day in the 30-month period. The results were then stratified into GDP days and non-GDP days, and averaged over the month in which the day occurred. In order to detect long-term trends, we smoothed the natural variance in the monthly IPE values by plotting a cumulative average (over all prior months) for each of the six metrics. The results are displayed in Figures 2 and 3 for SFO and EWR, respectively. The point on the ipe-6 GDP curve just above January 1998 in Figure 2, for instance, indicates that over all GDP days at SFO in that month, the average ipe-6 value (error) was about 29 minutes per flight. (Fights without a 6-hour history of departure predictions were excluded from the study.)

The monthly IPE averages for GDP days are substantially higher than for non-GDP days. This is to be expected, since the reassignment of arrival times by the FAA in a GDP leads to unpredictable variance in departure times and, subsequently, higher IPE values. Also, a GDP throws air carriers into a state of irregular operations that can have other adverse consequences affecting departure prediction.

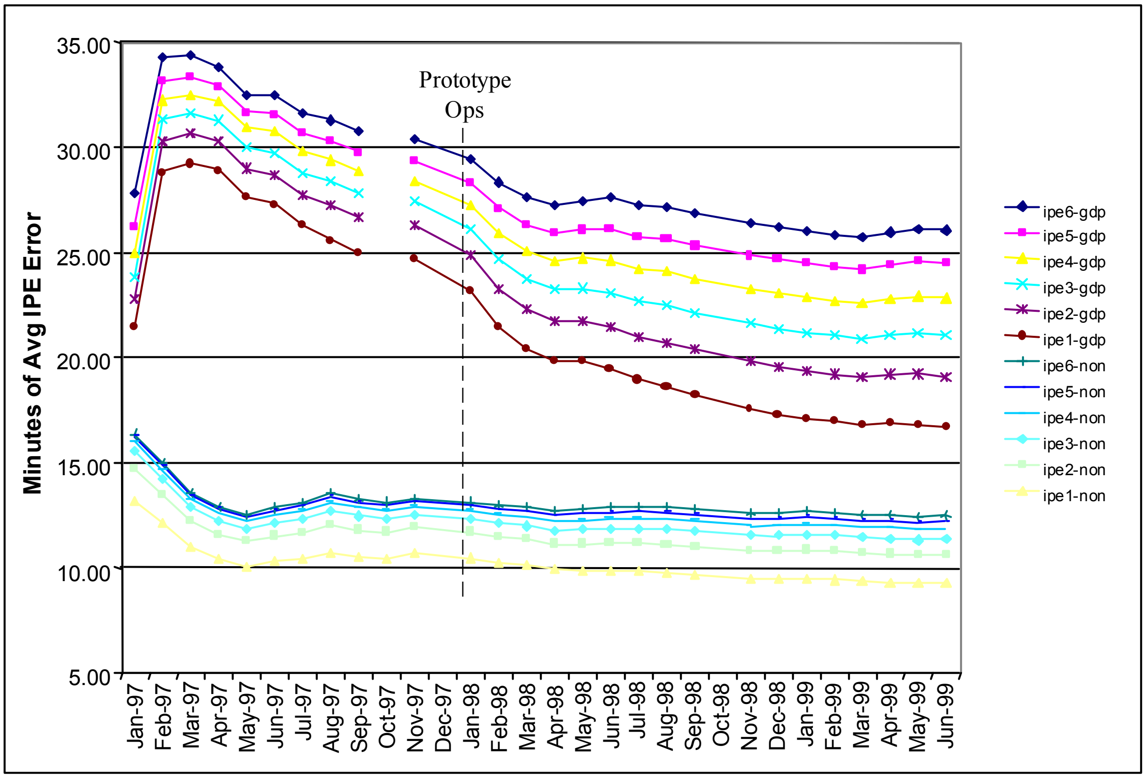

Figure 2. Cumulative Average of Monthly IPE Metric at SFO

The most notable feature of the charts is that on GDP days, when accurate flight data is most crucial, the average error in departure prediction dropped over most of the 30-month study, indicating an improvement in the accuracy of flight data. For instance, Figure 2 reveals that between January of 1998 (the inception of CDM) and June of 1999, the average ipe-6 departure error for GDP days at SFO dropped by 3.40 minutes per flight (5.58 minutes per flight for GDP days at EWR).[8] Average IPE values for non-GDP days at SFO and EWR have dropped as well. In fact, for both airports, the departure prediction error has been pushed below 15 minutes, the industry-wide standard for an on-time event.

Note that the positive (downward) trend actually began *before* the inception of prototype operations at SFO. We must acknowledge the possibility that this positive trend has occurred as a result of an inherent increase in the predictability of flight departures all across the NAS; this, however, seems highly unlikely. Also, The CDM participants began submitting flight data several months before prototype operations. A more plausible conclusion is that the improvement in departure predictions as shown by the IPE metric is a result of the commitment of the air carriers to submitting timely updates on flight departures.

[8] These are changes in the *cumulative* average since January of 1997. The non-cumulative average ipe-6 values for GDP days at SFO in Jan98 and Jun99 were: 28.82 minutes and 26.02 minutes, respectively (25.46 minutes and 22.33 minutes, respectively, for EWR).

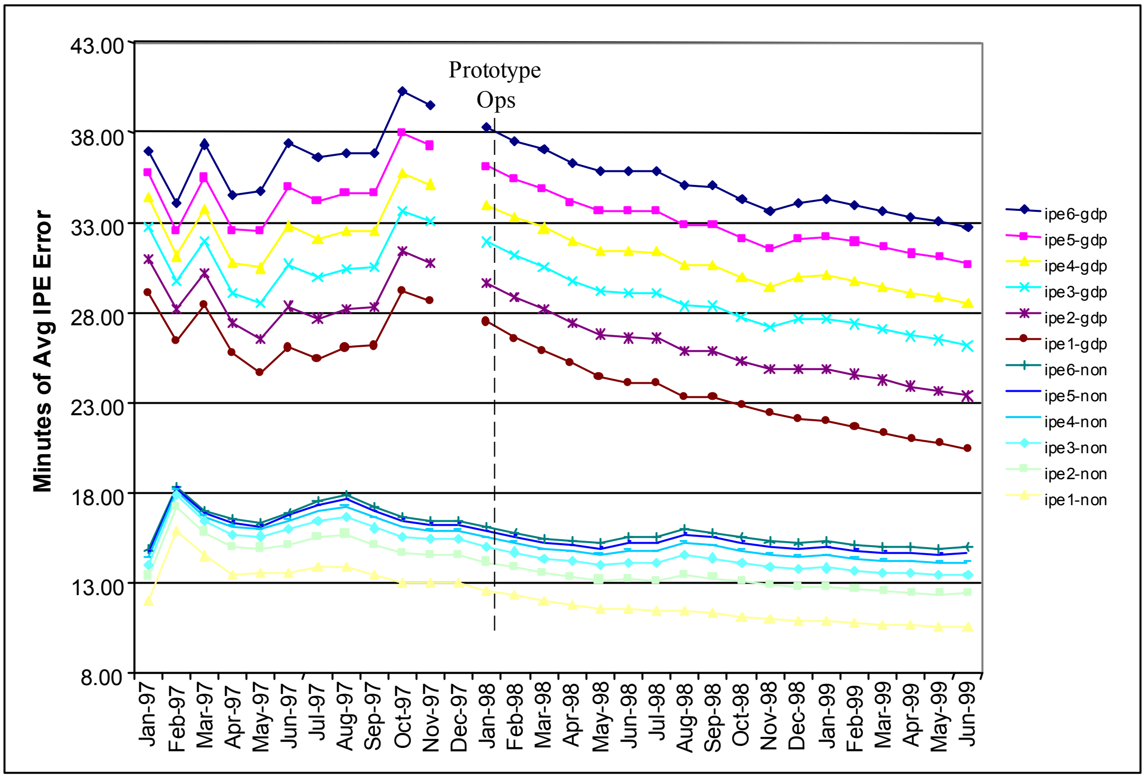

Figure 3. Cumulative Average of Monthly IPE Metric at EWR

There is no significance to the high variance at the left of the graph. This is a numerical consequence of forming a cumulative average over a series of fluctuating measurements: at first, the cumulative average is easily affected by single values but then stabilizes as it is formed over a larger number of values. Gaps in the lines indicate periods in which there were no GDPs or in which data was not available.

Note that all the metrics ipe-6, ipe-5, …, ipe-1, exhibit the a common pattern (for both GDP and non-GDP) but that the lower the tracking period, the lower the IPE value. For instance, the ipe-5 curve is essentially a downward shift of the ipe-6 curve, the ipe-4 curve is a downward shift of the ipe-5 curve, and so on. This indicates that, on average, departure prediction accuracy increases (has less error) as the departure of a flight approaches.

The IPE results were further stratified by airline (not shown for proprietary reasons). Oddly enough, the major air carriers participating in CDM tended to have the highest (worst) IPE values. We attribute this to the fact that these are the airlines with the most complex operational networks and therefore the most difficulty in predicting departure times; many of their flights are departing from locations with unpredictable delays due to congestion, weather, etc. We did observe, however, that the long-term results for each air carrier displayed a positive (downward) trend similar to that of the results aggregated over all air carriers.

It is unlikely that departure prediction error for GDP days will ever be reduced to the levels obtained on non-GDP days; both FAA and air carrier manipulations of departure schedules during a GDP will always introduce unpredictability in departure time predictions. Also, it is possible that improvements in departure time accuracy are approaching a limit, due to the fact that, in the aggregate, there will always be an inherent amount of uncertainty in prediction of aviation events.

2.4 Rate Control Index (RCI)

The rate control index (RCI) measures the flow of air traffic into an airport prior to any airborne holding that may take place and compares it to the targeted flow that was set by the traffic flow managers at the ATCSCC during a GDP. A single index, or percentage, is reported for the entire performance of the GDP on a single day. A higher score corresponds to better performance, meaning the flow of traffic into the airport closely matched the targeted pattern of traffic, both in quantity and in distribution. RCIs rarely go below 60 percent and usually hover around 90 percent. A perfect score of 100 percent is obtainable and has happened on several occasions. Since RCI does not, in and of itself, explain the good or bad execution of a GDP, GDPs with unusually high or low RCIs should be further analyzed for causality.

A GDP is an FAA action intended to take control of the arrival rate into an airport so that it matches airport acceptance rates (AARs) during a time horizon of concern. Since a GDP must be planned hours in advance, the AARs of the airport in question must be predicted based on weather forecasts. Because of the inherent inaccuracies in weather forecasting, AARs often turn out (in hindsight) to be higher or lower, respectively, than those targeted while planning a GDP. The purpose of this metric is to measure how closely the *execution* of a GDP matched the planned GDP, without holding the participants responsible for weather forecasts upon which the GDP was based.

The mathematics of the metric are rather complex and beyond the scope of this document. However, the basic mechanics of the metric can be summarized in the following idealized example. Suppose that a GDP is planned for hours 1, 2, 3, and 4. The first distribution in Figure 4 shows the number of flights that are planned to arrive into the terminal space of an airport (but not necessarily land) for each hour of the GDP. This is the *planned* or *desired* distribution of flights. The second distribution in Figure 7, the *realized* distribution, shows the actual number of arrivals at the terminal space, recorded after the fact. (More precisely, this is the number of arrivals that would have taken place in the absence of airborne holding and given sufficient arrival capacity.)

The RCI metric computes the minimum amount of (fictitious) flight movement that would be necessary to revert the realized distribution to the desired distribution by shifting flights between the columns of the realized distribution (i.e., move 2 flights hour 1 to hour 2, move 2 flights from hour 3 to hour 4 and so on). This number is 8 flight-hours. The worst-case scenario is 330 flight-hours (details of computation omitted) which implies that the realized distribution is off from the desired distribution by $8/330 \times 100$ percent = 2.42 percent. So the GDP receives an RCI of $100 - 2.42 = 97.58$ percent. A similar computation shows that, if the realized distribution had been noticeably worse, as in the third distribution of Figure 4, then the GDP would have received an RCI of only 75.75 percent.

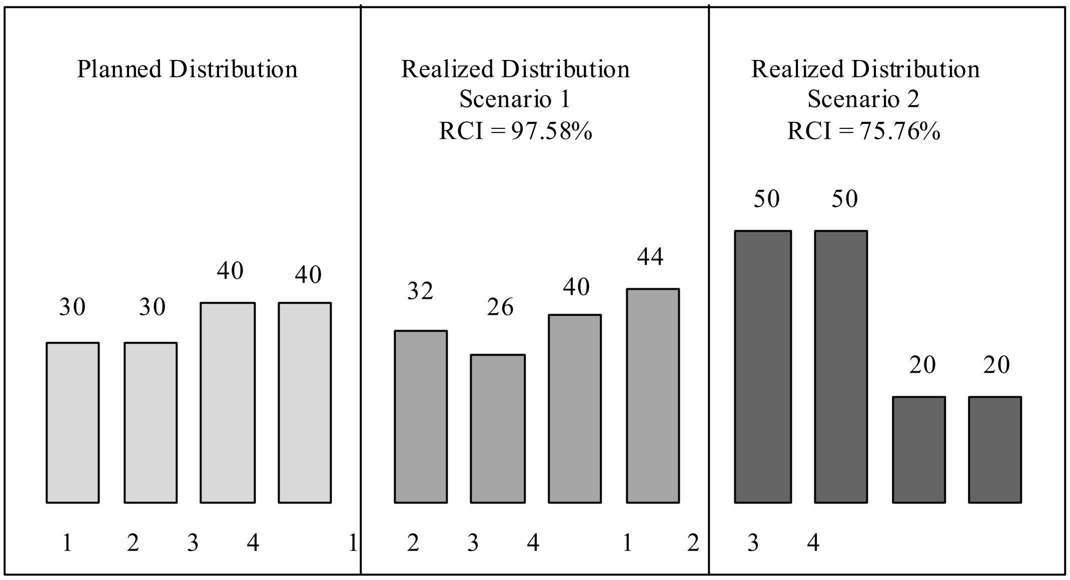

Figure 4. Rate Control Index Distribution

RCI will be used to compare pre-CDM performance against post CDM performance, as well as to spot long-term trends in GDP performance. In particular, RCI is adept at flagging poor performance GDPs and at highlighting those GDPs that ran particularly smoothly. Computation of the RCI index requires the following data for each flight arriving at a GDP airport during the GDP time interval: (1) an arrival time and (2) a sequence of en route arrival predictions. From this, an hourly count can be formed that reflects the flow of traffic into the airport that would have occurred, had there been sufficient capacity and no airborne holding. This distribution of traffic is weighed against the distribution formed from the final (last revised) planned AAR for each hour of the GDP. The flight data is readily available from archived ETMS data. The PAAR data has been recorded from ATCSCC logs and archived by Metron.

The RCI captures many elements of GDP performance including EDCT compliance, cancellation notices, delays, and forecasted demand. Since (by design) RCI does not capture throughput at an airport, nor the ability of local controllers to process an arrival stream, these factors should be measured in separate metrics.

Figures 5 and 6 show the trend of the RCI metric for SFO and EWR, respectively. The results were recorded over the 30-month period from January 1997 to March 1999. This captures a full year of data prior to the implementation of CDM prototype operations in January of 1998. Note that there is much greater variance in the monthly RCI value at EWR than at SFO. (See the lines with round icons.) Also, the monthly RCI values are generally lower at EWR than at SFO. Each of these is likely due to the unpredictable nature of East Coast traffic and the complexity of the airspace surrounding the New York City area (it lies on the boundary of more than one Center).

There appears to be a seasonal drop in the RCI value at each airport. For SFO, the drop occurs in the winter months; for EWR, the drop occurs in the summer months. These may be due to weather conditions. The drops were most pronounced in 1998, when the effects of El Nino were at their greatest. Although the RCI metric screens out effects of weather conditions immediately at the airport, traffic flow into the airport could be affected by weather conditions elsewhere in the NAS. For instance, delays on in the Northeast make traffic flow into SFO sluggish or erratic, thus lowering the RCI value.

We smoothed out the variance of the monthly points by computing a moving average over four months (see the lines with the square icons in Figures 5 and 6). At SFO, a slight upward trend is visible. Further smoothing was obtained by computing a cumulative average over all months since January of 1997 (see the lines with the triangular icons). The cumulative average serves as our best evaluation of a long-term trend and reveals a slight upward trend at each of the airports, indicating an improvement. There are two time periods of concern: the pre-CDM period (January 1997 – January 1998) and the post-CDM period (January 1998 – March 1999). At SFO, there was a small improvement over each period (from 91.42 to 92.33, then from 92.33 to 92.75). At EWR, there was a slight decrease in performance in the pre-CDM period (from 89.13 to 86.56), then an increase in the post-CDM period (from 86.56 to 87.91).

So, we conclude that at both SFO and EWR, there has been on average definite (but not large) improvement in GDP-controlled traffic flow while CDM has been in operation. More improvement has been seen at EWR than at SFO, and yet overall performance is lower at EWR.

We caution that these long-term trends in the RCI should not be over-interpreted. First, the results will vary with the method of computation. Since the number of ground delay programs varies with the month, these results do not give equal weight to ground delay programs. Different results are obtained when equal weight is given to each program (i.e., averaging over all prior ground delay programs). The method we have adopted screens out some of the seasonal effects of weather.[9] Also, the computation of the metric at the terminal space, as opposed to on the runway,[10] relies upon modeling post facto the airborne holding that each flight endured during the ground delay program. Other models may produce different results, particularly as to how much airborne holding was endured by an individual flight. We are confident, however, that in the aggregate the same long-term trends will result.

[9] The legitimacy of our method has been confirmed by averaging over each season of the year. Results similar to what we have presented here are obtained.
[10] The RCI metric can be used to meter the flow of traffic onto the runways against the desired flow. This is partly how the ATCSCC measures the success of a program. However, this allows the ability of local controllers to land aircraft and the effects of local weather to enter into the results.

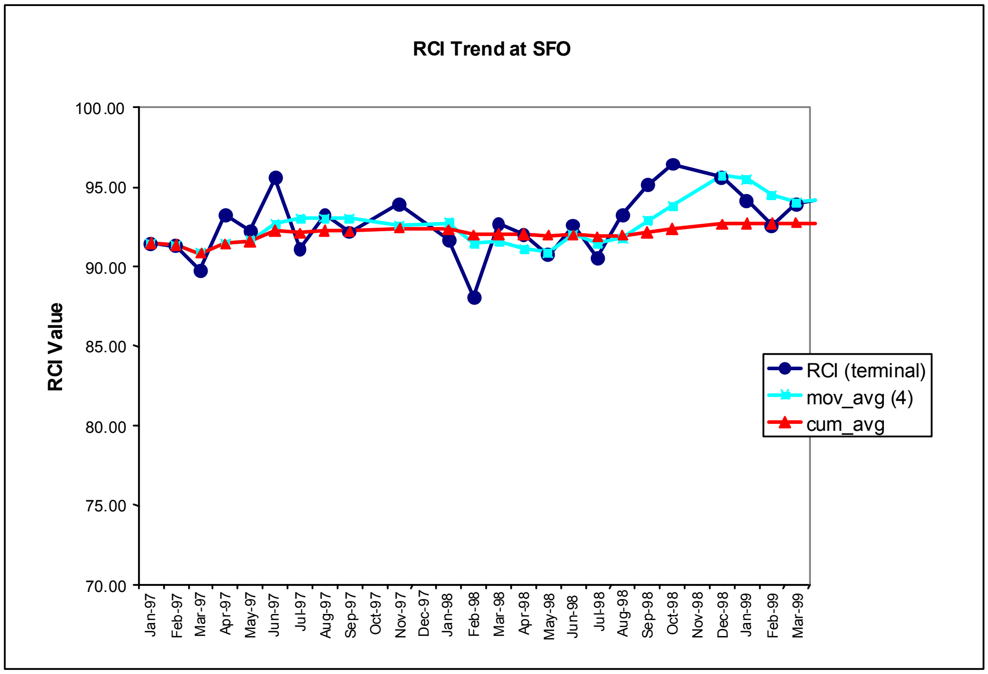

Figure 5: Cumulative Average of the RCI Metric at SFO

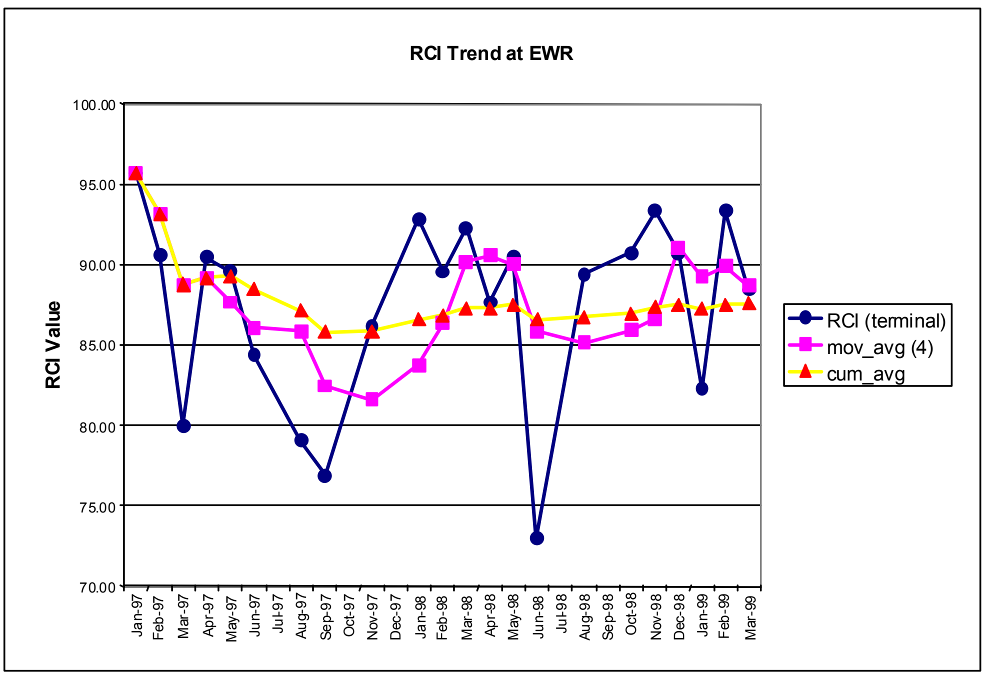

Figure 6: Cumulative Average of the RCI Metric at EWR

2.5 GDPs Canceled Near Start

This type of GDP cancellation occurs within the time frame from 30 minutes prior to the start time of a GDP to 30 minutes after the start time of the GDP. In this context start time of a GDP refers to the time the first controlled flight is scheduled to land. Since ground delays are assigned and taken before flights take off, many fights will have absorbed delays well in advance of the start time of the GDP. If a GDP is canceled at or near its start, this is an indication that there was no need for the GDP. Thus, all the (sometimes substantial) delays that have been absorbed prior to the start of the canceled GDP were found to be unnecessary. If there is a need for a cancellation of a GDP, then ideally it should be done in timely enough fashion to allow airlines to recover delay caused by the institution of the GDP. Under CDM specialists have much more accurate demand information to plan with. Furthermore, they have the opportunity to delay the institution of a GDP until the *critical decision point*. Since they can delay it for a much longer period of time than during the *Pre-CDM* phase, they have the option to wait until better and more accurate weather and demand information is available. Thus, decisions to institute GDPs are based on superior data quality and accuracy and should generally result in fewer cancellations near GDP start times.

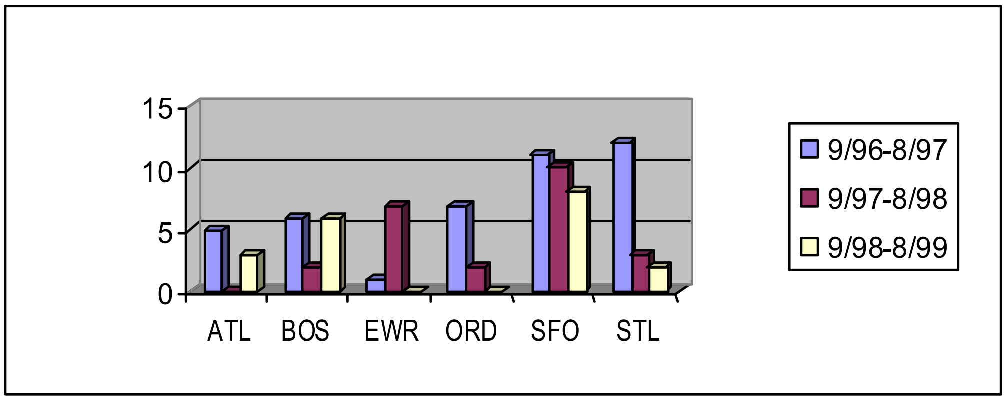

Figure 7. Percentage of GDP Cancellations Near Start

As illustrated in Figure 7, percentage of GDP cancellations near start indeed decrease between the *pre-CDM* and the *CDM* phases, although there is a consistent downward trend at ORD, SFO and STL through all phases.

There has been a remarkable improvement at STL in the percentage of canceled GDPs near the start time. A likely reason that may contribute to this trend at STL is the superior data quality of two major airlines that utilize this airport heavily. This caliber of data quality can be attributed to the use of daily download. In addition, these airlines have provided positive feedback on FSM and procedures under CDM.

2.6 Anecdotal Benefits of GDP-E

One of the efforts in accessing the benefits of CDM is collecting anecdotal evidence from both the user community and the command center. The approach taken was different for each of the user groups. For the air carriers, a questionnaire was sent by email listing ten questions to be answered. For the command center, the primary source of anecdotal evidence was taken from the ATCSCC specialists' logs, mostly in the GDP critiques.

Four of the carriers responded to the questionnaire and their comments along with those provided by the ATCSCC are included in Appendix C.

This Page is Intentionally Blank

3.0 Compression

3.1 Compression Benefits and Methodology

Compression savings by airport have been tracked since the beginning of prototype operations on January 20, 1998. These compression benefits can also be displayed cumulatively over time. Figure 8 shows how the compression benefits have increased since the beginning of prototype operations. Three key events are highlighted on the graph; the start of all airports on September 8, 1998, the severe weather season over the winter, and the date in which the slot allocation algorithm was changed from RBS to RBS++ (compression included). You can see the how the slope of the line increased at the "all airports" mark, and then increased again for the snow season. The more GDPs that are run (and the more the subsequent compression cycles), the steeper the slope of the line will be. It has remained fairly steady since March 18, 1999.

The dates in which certain milestones were reached in compression benefits are also presented below. The first column contains the milestones based on accumulating savings from the beginning of prototype operations. The second column is based on benefits starting at the time operations expanded to all airports. There have been a total of 3,165,925 cumulative minutes reduced due to compression since the start of prototype operations.

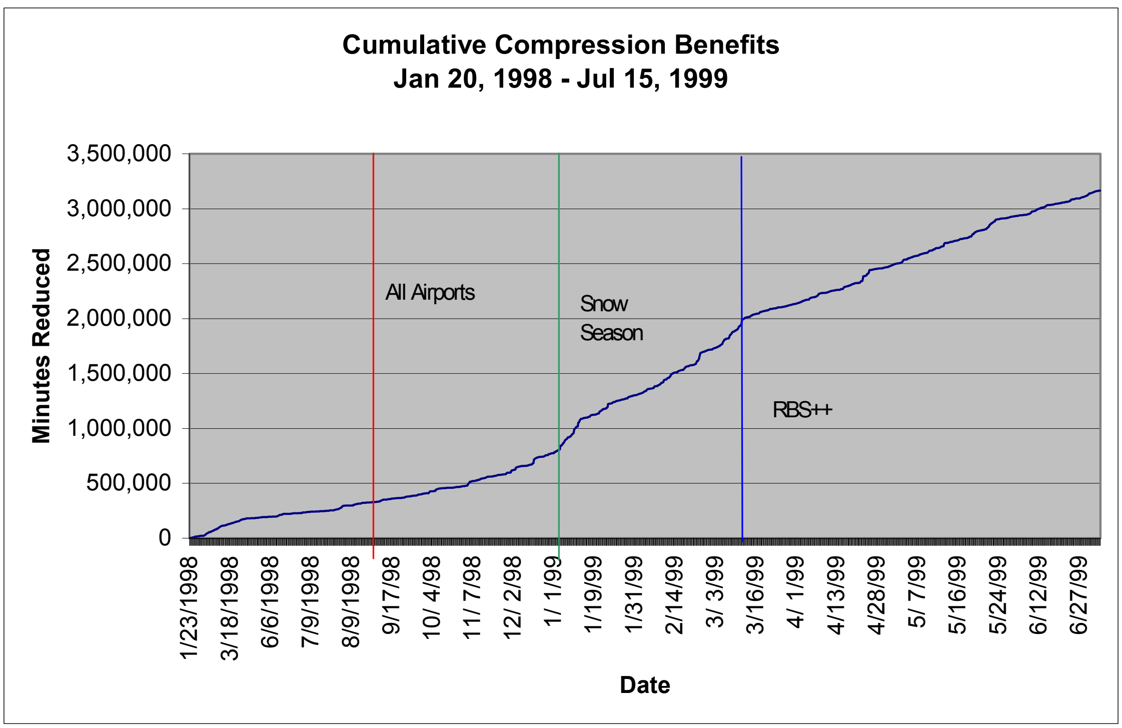

Figure 8. Cumulative Compression Savings Since Start of Prototype Operations

Starting in September 8, 1998, when prototype operations expanded to all airports, compression benefits were tracked by CDM-participating carriers. During that time period, there were a total of 1,091 compression cycles, some just compression and some being the compression portion of RBS++.

Figure 9 shows the total minutes reduced by carrier, sorted from highest to lowest. It illustrates minutes reduced by airport and airline. The "GA/M" category includes all general aviation and military flights. The "Other" category includes all non-CDM-participating carriers. Starting on October 13, 1998, a policy change was made at the ATCSCC to include non-CDM-participating carriers in the compression process. Prior to that, their flights were never moved up by compression.

Most of the carriers listed in Figure 9 have been CDM-participants since the beginning of prototype operations. The exceptions are Midwest Express (MEP), and America West (AWE), who became participants on February 1, 1999, and Federal Express (FDX), which became a participant on February 16, 1999.

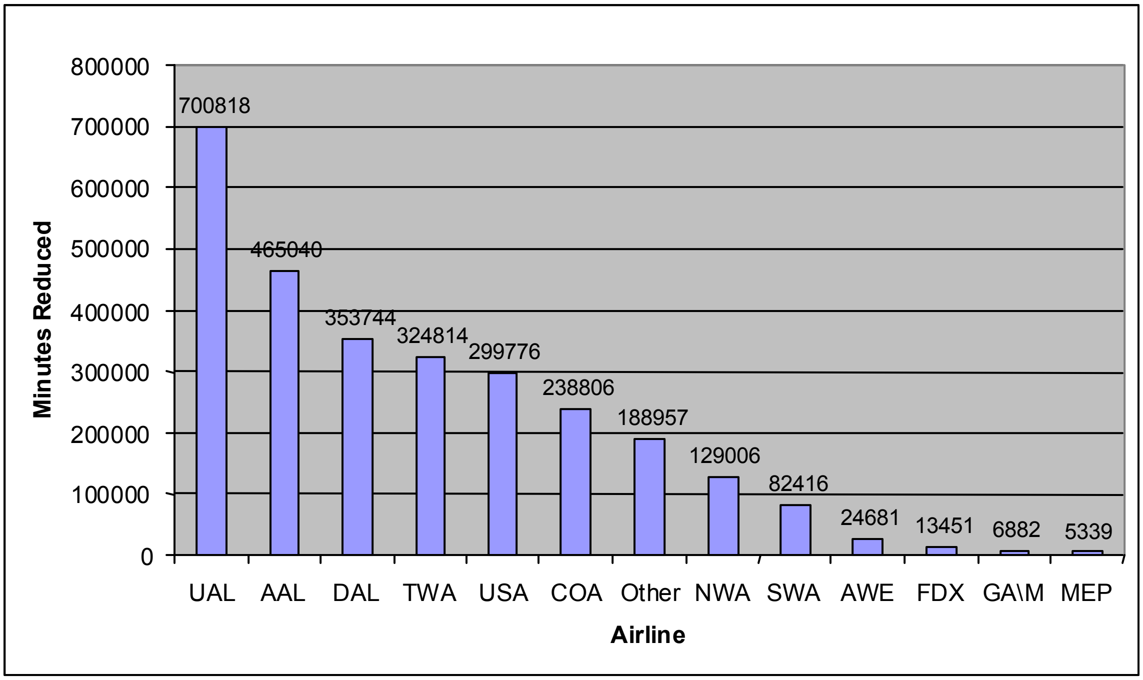

Figure 9. Minutes Reduced Due to Compression/RBS++ by Airline (September 8, 1998 – July 15, 1999)

We have found that the amount of minutes reduced for each carrier is highly dependent on a number of factors:

1. The size of the carrier (percent of all traffic).
2. Location of major hub airports and how many GDPs were run at those airports.
3. Number of cancellations/open slots in compressed.
4. GDPs - bridge-only status.

In this graph, UAL has by far the highest number of minutes reduced. This is due primarily to #1 and #2 listed above. AAL's and DAL's results are also for the same reason.

TWA's results, on the other hand, are due primarily to #3. They rely on the compression process instead of doing internal substitutions to fill in the open slots created by their cancellations.

The *Other* category has fairly good results compared with some of the CDM-participants, but this is due to #1 above. The "Other" flights comprise a large percentage of the traffic. Actually, as shown in Figure 10, the only carriers with a higher traffic percentage are UAL, AAL, and DAL. The airlines with the lowest traffic percentage are AWE, FDX, and MEP however their savings only started accumulating in February 1999.

Figure 10 presents two percentage values are shown for each carrier. The first number, percent change, is the percentage of total delay reduced on average due to compression. The second number, traffic percent, is the percent of the total traffic which is operated by that carrier.

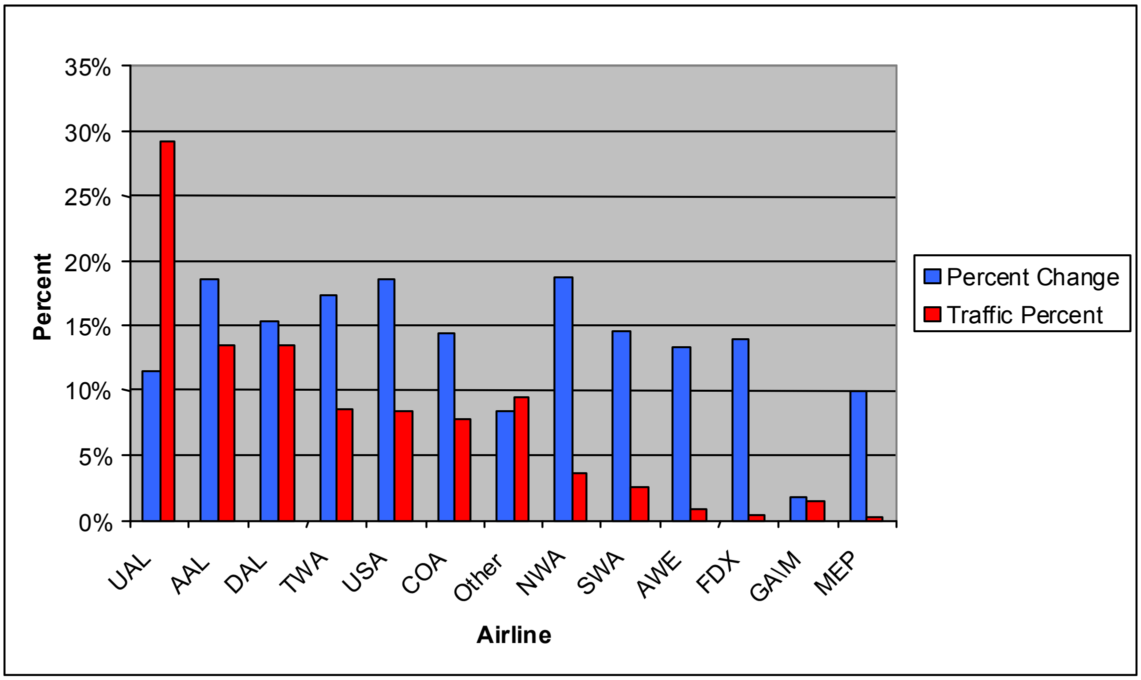

Figure 10. Percent Reduction vs Percent Traffic by Airline (September 8, 1998 – July 15, 1999)

The carriers are still sorted by total minutes of delay reduced but as can be seen, the percent reduction does not correspond with the total minutes reduced. Actually, UAL has a much lower percentage than many of the carriers. This is due primarily to their bridge-only status. Since they are a bridge-only in seven major hub airports, many of which frequently have GDPs, their compression benefits are lower than they would have been without the bridge-only status. UAL prefers to utilize the substitution process in moving flights up to fill open slots due to cancellations.

An interesting thing to note is that the amount of traffic a carrier has does not affect the percent savings achieved through compression. All of the carriers with the smallest amount of traffic all have significant percent compression savings. This proves the concept of how compression can significantly help the smaller carriers when the substitution process cannot.

The two data points that stand out with smaller percent reductions are *Other* and *GA/M*. The *Other* flights were not included in compression until October 13, 1998, and even then, they receive a lower priority in compression than the CDM-participants. The fact that the *Other* percentage is so much lower than the other carriers shows the benefits to becoming a CDM-participant.

This report provides a detailed explanation on the methodology of how compression benefits are calculated.

Introduction and Analysis Methodology

Compression is a process used to ensure that no valuable arrival slots at an airport go unused during a ground delay program. This process can be run multiple times during the course of a ground delay program, as need dictates. The algorithm identifies open arrival slots due to flight cancellations and delays. It then moves other flights up, reducing their delays, to fill the vacated slots. Compression always attempts to fill an open slot by moving up a flight which belongs to the same carrier as the open slot. If that is not possible, it then tries to find a CDM-participating carrier that can benefit from the slot. Otherwise, the slot is made available to all flights.

These analyses track the number of minutes reduced to flights' estimated time of arrival every time a compression is run at the ATCSCC. These numbers have been recorded since the beginning of prototype operations on January 20, 1998. Since September 8, 1998, when CDM began operation at all airports, the numbers have been tracked in greater detail, including the delay reductions broken out by carrier. Starting on March 18, 1999, the delay reductions have also been tracked by individual flight.

For this particular metric, there is no available baseline, or pre-CDM data. Compression is a new concept introduced through CDM. Some of these reductions due to compression could have been achieved by the carriers prior to CDM via the substitution process. This portion of the compression benefits has been quantified, but there is still a significant amount of benefits that could have been achieved only through compression.

Cumulative Scheduled vs. Actual Scheduled Compression Benefits

The planned compression savings tracks the cumulative amount of delay reduced for each flight over every compression cycle. Not all of these compression savings are actually realized. This can be due to one of three major reasons.

1. Change in the scheduled delay: a flight's controlled time is revised via a GDP revision prior to departure, thus overriding the earlier compression benefits. Taking this into account results in actual scheduled delay reduction, as opposed to cumulative scheduled delay reduction.
2. Cancellation: A flight which was included in a compression cycle is never flown.
3. Other NAS impacts: a flight experiences a delay in departure, enroute airborne delays, or airborne holding at the arrival airport.

The third factor is extremely difficult to measure. It is hard to correlate airborne delay with a compression cycle. In this analysis, only the first and second reasons are addressed.

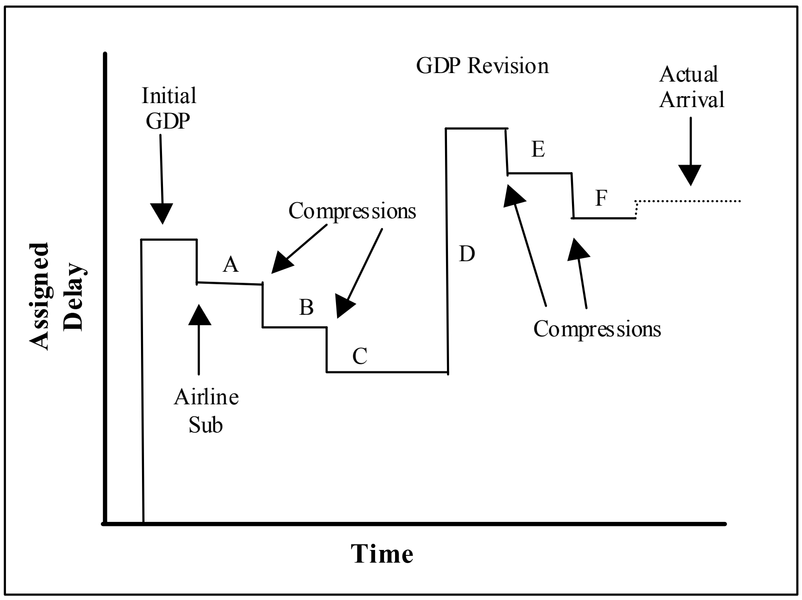

Figure 11. Graphical Representation of Compression Calculation

The following example illustrates the methodology used for a fictitious flight:

In this example, the delay additions/reductions illustrated on the graph are:

A = -15	D = +45
B = -10	E = -15
C = -8	F = -10

Value 1: Cumulative scheduled delay reduction: B+C+E+F = -43

Value 2: Actual scheduled delay reduction: E+F = -25 (58 percent of value 1)

Value 2 does not include the benefits achieved from the first 2 compression cycles since the flight was revised after that prior to departure.

The methodology used in this analysis is as follows:

1. All flights included in a compression cycle which were later included in a revision prior to departure were removed.
2. All flights which were later canceled after being included in compression are removed.

This analysis looked at all of the compression cycles over a 3 month period; April, May and June of 1999. This includes 264 GDPs, 415 RBS++ events, 258 compression cycles, and 93,169 flights affected. The results were compiled by month. The delay reductions resulting from the above methodology were calculated as a percent of the cumulative scheduled delay reductions. The results are shown in the following table.

Table 2. Compression Cycle Data Comparison

Month	Cumulative Compression Reduction	Actual Compression Reduction	Actual as a percentage of Cumulative
April	329,807	147,756	45%
May	469,861	258,206	55%
June	188,506	83,255	44%
Total	**988,174**	**489,217**	**50%**

Using the data in Table 2, our best estimate of actual scheduled compression benefits is about one-half of cumulative compression benefits.

The actual scheduled compression benefits can also be graphed as a percentage of the cumulative scheduled compression benefits, as shown on the following graph. The average percent over the 3-month period is 50 percent. The following graph plots the cumulative scheduled compression benefits by month on top, and the actual scheduled compression benefits on the bottom.

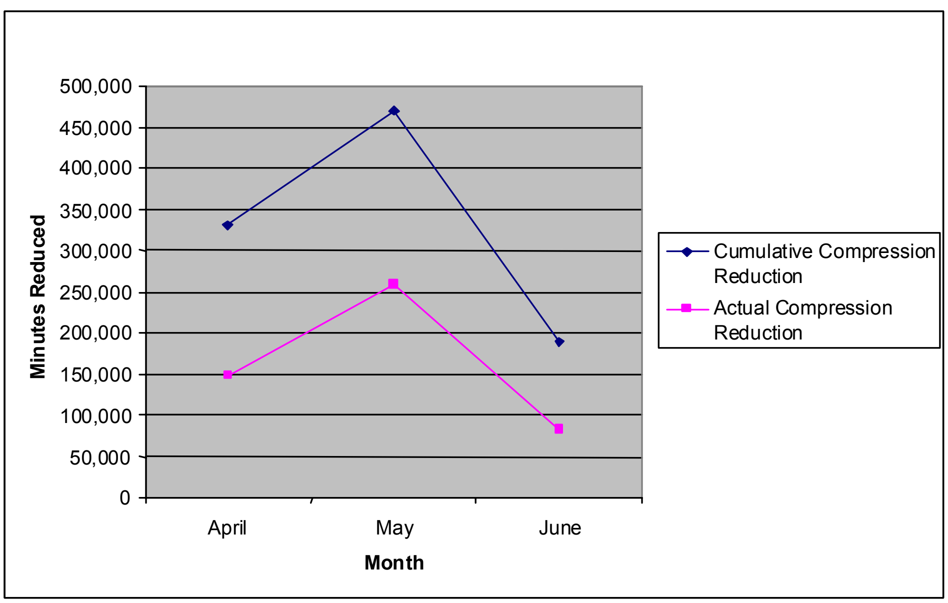

Figure 12. Cumulative vs. Actual Compression Benefits

Gross Compression vs. Net Compression

Since an open slot is always first made available to the carrier which owns the slot, some of the delay reductions achieved from compression are due to moving flights up into slots they have vacated due to flight cancellations. Carriers have already had that capability provided to them prior to CDM. It is called the substitution process. Any carrier can cancel a flight that has been assigned an arrival slot in a GDP and move another one of their flights up to fill the open slot.

The purpose of this analysis was to identify what component of the benefits achieved through compression could have been achieved by the carriers via the substitution process. Utilizing FSM in historical mode, compression can be re-modeled and the results analyzed. FSM also includes a substitution algorithm which models how delays can be reduced by the carriers using the substitution process. The methodology used for this study was the following:

1. Utilizing FSM in historical mode, set the time to match the model time of the compression.
2. Use the "Subs" algorithm in FSM to model the delay reductions which can be achieve from substitution.
3. Do not reset the data.
4. Run the "Compression" algorithm in FSM using the same parameters that were used initially.
5. The minutes reduced due to this compression are the net compression benefits.
6. The difference between the total reduction and the net compression reduction is the substitution, or gross compression, reduction.

By running the substitution algorithm prior to running compression, all benefits which could have been achieved by the carriers alone using the substitution process are removed. The resulting benefits are delay savings that could never have been achieved prior to CDM/GDPE.

The methodology for identifying net compression savings is conservative. This is due to several reasons. Even though a carrier could have achieved the benefits using the substitution process, some carriers have not implemented this capability and are not able to carry out substitutions. Also, even though a carrier may use the substitution process, it can be difficult to recognize every opportunity for a sub. In the past, arrival slots frequently went unused because the carrier that owned that slot did not substitute one of their flights up to fill it. It can become especially difficult during the more severe weather events. The more a carrier's schedule is being disrupted, the harder it is to monitor substitution opportunities. In these cases, compression has provided a tool for the ATCSCC to do their substitutions for them, the results being overall delay reductions across all of their flights.

This analytical methodology was applied over all compression cycles between September 8, 1998 and March 17, 1999. This included 585 compressions over 246 ground delay programs at 15 airports. The average breakout between subs and pure compression over all of these compression cycles was 66 percent of the compression benefits could have been achieved theoretically via the substitution process and 34 percent could have been achieved only through compression. The results are also displayed in Figure 13.

The airports are sorted from highest minutes of compression savings to lowest. The total height of each bar is the total percent reduction due to compression. Each bar is broken into two sections; the top section for net compression savings and the bottom section for gross compression (or substitution) savings.

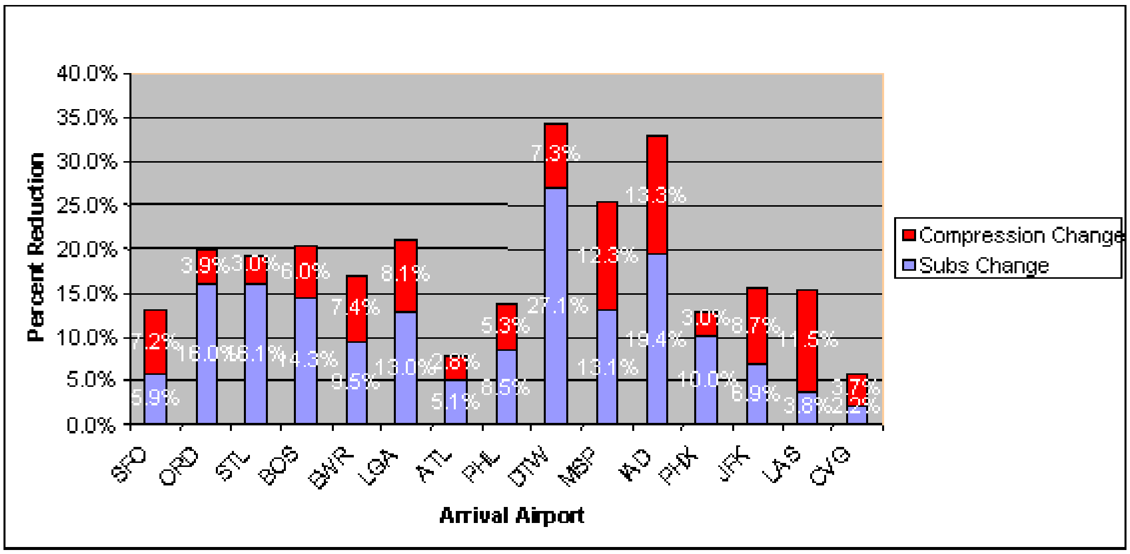

Figure 13. Percent of Delay Reduction Due to Compression

Two airports on the opposite end of the spectrum in this graph are STL and SFO. STL has a low percent of net compression savings, 15.7 percent. This is due to the fact that STL traffic is primarily represented by TWA. TWA prefers to let the compression process do their substitutions for them. Thus a large percentage of compression savings at STL could have been achieved by TWA utilizing the substitution process.

SFO on the other hand has large percent of net compression savings, 55 percent. SFO also has a predominant carrier, UAL, but UAL does not have as large a percent of traffic in SFO (approximately 56 percent) as TWA has in STL (approximately 71 percent). SFO has a wide range of carriers with small percents of the traffic. These carriers rely on the compression process to take advantage of their canceled open slots, where substitution is unable to do so.

LAS, CVG, and JFK also have a large percentage of net compression savings, 75.4, 62.9, and 55.5 percent, respectively. But these numbers are based on a small data set, two, nine, and eight compression cycles for each of the airports.

Over all airports the average savings was approximately 1/3 due to pure compression and 2/3 savings due to theoretical substitution. We believe that a big enough data set was used to establish this trend. Thus, breaking out compression benefits between net compression and gross compression was not continued after March 17, 1999.

Delay Reduction Distributions by Flight

The compression benefits analyses discussed so far only address the cumulative minutes reduced by airport and/or carrier. None of these analyses have looked at the minutes reduced on an individual flight level. Are only a few flights receiving large reductions while the majority receives no reductions, or are the delay savings fairly evenly distributed among the flights? This analysis looked at the delay reductions for each individual flight to see how they are distributed.

The time period included in this analysis is from March 18 – July 15, 1999. This includes 521 compression cycles over 190 GDPs. A total of 96,603 flights were captured in these compression cycles, accounting for a total of 1,105,012 minutes reduced. Figure 14 displays the delay distributions in a histogram. It shows the number of flights which fall into each of the 15 minute buckets of delay reduced.

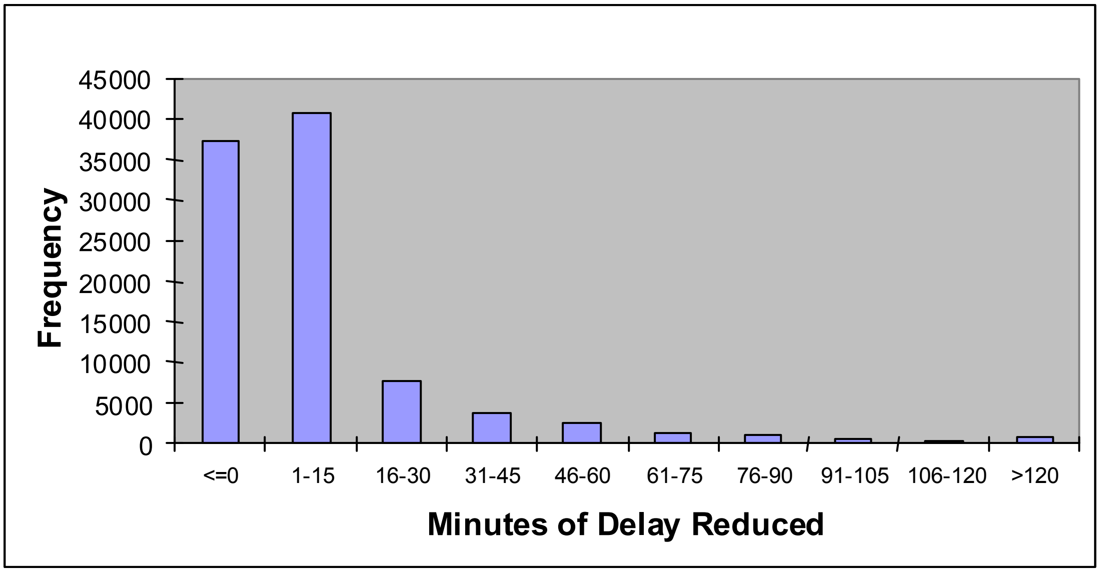

Figure 14. Delay Reduction Due to Compression (March 18 – July 15, 1999)

On average, 39 percent of the flights receive no reduction due to compression, and 61 percent are receiving at least 1 minute of delay reduced. Some key percentage measures are:

Reductions 15 minutes or greater: 19 percent
Reductions 30 minutes or greater: 11 percent
Reductions 45 minutes or greater: 7 percent
Reductions 60 minutes or greater: 5 percent

The average reduction per flight is 11.5 minutes, with a median reduction of 2 minutes.

Assumptions

The planned delay reductions due to compression are being archived for every flight included in every compression run at the ATCSCC, including the compression portion of RBS++. Every time a GDP or compression is run at the ATCSCC, a report is generated which lists the changes in delay due to that action by both carrier and individual flight. These reports are being loaded into a database on a weekly basis. Several reports and queries have been created which will generate the compression benefits for any time period specified.

The types of compression benefits analyses included in this analysis are:

1. Savings by airport (including minutes and percent reduced): Jan. 20, 1998 – Jul. 15, 1999
2. Savings by airport and carrier (including minutes and percent reduced): Sep. 8, 1998 – Jul. 15, 1999
3. Benefits due to pure compression vs. theoretical substitution: Sep. 8, 1998 – Mar. 17, 1999
4. Delay reduction distributions by flight: Mar. 18 – Jul 15, 1999
5. Cumulative scheduled vs. actual scheduled compression benefits: Apr. 1 – Jun. 30, 1999

Anecdotal Evidence of Compression Benefits

A number of positive testimonials about compression have been received from CDM-participating carriers. A few of them are listed here.

Midwest Express, MEP, reports “We have estimated for the first 2 1/2 months we saved approximately $50,000 + due to reduced delays because of compression. This figure was derived using a $25/delay minute times the number of minutes our gate-holds were reduced by compression.”

Business Express, GAA, reports “There have been GDPs (no specifics dates) where we were able to sub up our flights to a point where we could actually manage the delays in the GDP, and after a compression, have noticed even more reductions (bonus!)”

Transworld Airlines (TWA) reports; “When ground delay program advisories are sent we build program to see what the impact may be before EDCT times are sent. We also do the same with compression which has given us some early heads up on reduced times.” He later states “Compression especially has been helpful at stations like EWR/DFW/ATL/ORD where we are small players.” And finally he states “One quick story. We had a program for EWR about a week ago and we had approximately a 90-minute delay. The command center advised they were going to run compression at a specific time. I quickly ran compression on our end and our delay was reduced to down to 20 minutes. I repeated the drill to make sure I was right and then advised the station to get flight ready to go that the delay would be reduced. When compression was run I missed time by 4 minutes.”

A complete list of airline testimonials can be found in a separate paper, CDM Anecdotal Evidence for FFP1 Performance Metrics.

Summary

The compression process, which was introduced by the GDP enhancements of CDM, has proven to provide substantial benefits to the user community, as seen by these analyses. But what these numbers do not show is the benefit also provided to the FAA/NAS. Compression provides the ATCSCC with a tool which helps provide a smooth arrival rate into an airport, without wasting valuable arrival resources. As a result of CDM, the ATCSCC has more information about cancellations and delays, and can use compression to fill those slots which are not utilized through the substitution process.

The two main questions about these compression benefits analyses have been:

1. How much of these benefits could have been achieved through substitution?
2. How much of these savings are ever actually realized?

Both of these questions are answered in this paper. These conservative methods of calculating discounting both the savings which could have been achieved via the substitution process and discounting savings which were never realized still results in significant savings in delay minutes to the industry.

3.2 Equity

GDP-E introduced a new process for assigning flights to arrival slots during a ground delay program. This algorithm is called RBS, or Ration-by-Schedule. The carrier community along with the FAA worked together to define this process so that it would be fair and equitable to all concerned. The two major changes to the slot allocation process in RBS are:

1. Flights are sorted and prioritized based on their original scheduled times, as opposed to their current ETA, as was done using Grover Jack. This ensures that no flight will be penalized by reporting an airline delay prior to a GDP.
2. Canceled flights are allocated arrival slots, not just non-canceled flights. This ensures that a carrier will not be penalized by receiving one less slot by reporting a flight cancellation prior to a GDP.

In a completely fair and equitable GDP, each carrier should receive a percentage of the total assigned delay which is equal to the percentage of the total traffic for their carrier. This can be stated for one carrier as follows:

(Delay assigned to carrier / Total delay over all users) / (Carrier's Flights / Total Flights) = 1

But there can be a few reasons as to why a carrier's equity metric would not be equal to 1.

1. Schedule composition: The location of a flight's original schedule time compared to the other flights in the GDP has a big impact in the delay which will be assigned. In many cases, the earlier a flight's original scheduled arrival time is, the less delay that flight will receive. In other cases, the delays rise to a peak during the GDP duration and then start to decline. A flight with an original scheduled time near the beginning or the end will receive a smaller than average delay.
2. Short haul vs. long haul traffic: The majority of GDPs do not include all centers in the program. Most will include a subset of the centers which are the closest to the airport, and will exempt the centers that are farther away. This means that a carrier that has more short-haul flights as opposed to long-haul will be assigned higher total delays because less of their flights will be exempted due to departure center.
3. Effect of compression: GDPs are now run using the RBS++ algorithm, which adds the compression process to the RBS rationing scheme. This means that the delays assigned to flights in a GDP will be affected by the results of compression. If a carrier has sent in more cancellations and/or delays than other carriers prior to the GDP, they will probably get more delay reduced due to compression. It will look like this carrier is getting a lower equity metric than the other carriers, but this is due to his cancellation strategy.

This analysis presents the results of analyzing the equity metrics over the participating CDM member carriers. For each carrier, four separate equity metrics have been calculated. The reason for calculating the equity metric in four different ways is to address the issues stated above in #2 and #3. The effects of both the centers included in a GDP and the compression process of RBS++ can be removed. This leaves only the effect of #1, schedule composition, as the factor affecting the equity metric.

The four different equity metrics calculated in this analysis are:

1. Equity over non-exempt flights, without compression: This is the metric that discounts the effects of center inclusion and compression. The delay is calculated by using the delay assigned after the RBS portion of RBS++ and before the compression portion of RBS++. The number of flights is calculated over the non-exempt flights only.
2. Equity over all flights, without compression: This metric discounts the effects of compression only. The number of flights is calculated over all controlled flights, not just the non-exempt flights.
3. Equity over non-exempt flights, with compression: This metric discounts the effects of center inclusion only. The number of flights is calculated over the non-exempt flights only.
4. Equity over all flights, with compression: This metric does not discount any of the effects due to compression or center inclusion. The delay is calculated using the results after RBS++, which includes compression, and the flight counts are over all controlled flights, not just the non-exempt flights.

Because the first equity discounts the effects of center inclusion and compression, this is the metric that should be closest to 1.0 for all carriers. The variation in this metric for each of the carriers is most likely the effect of their schedule composition. The argument is that this metric is the one that best represents carrier equity of assigned delays.

This analysis calculates the equity metrics by carrier. It includes all GDPs from March 18 – July 31, 1999, a total of 230 programs with RBS++ run 453 times. There are almost 140,000 flights captured in these 453 runs of RBS++, 83,000 of which were non-exempt flights.

Table 3 shows the results of all four equity metrics for each carrier. GA/Military flights and Non-CDM participating carriers are included as two of the categories, "GA/M" and "Other" respectively. The carriers included are those carriers who are CDM participants. The subcarriers flights of each of the major carriers listed are included in the analysis. A metric of less than one implies that the carrier received less than their fair share of the delay, while a measure of greater than one implies that the carrier received more than their fair share of the delay. The carriers are sorted by the first equity metric, best to worse, or smallest to greatest.

Table 3. Compression and Equity by Airline

	Without Compression:		With Compression:	
Airline	Equity Over Non-exempt Flights	Equity Over All Flights	Equity Over Non-exempt Flights	Equity Over All Flights
DAL	0.90	0.98	0.86	0.93
AWE	0.93	0.93	0.97	0.95
FDX	0.98	1.16	1.20	1.14
SWA	0.98	1.35	0.99	1.21
USA	0.98	1.14	0.89	1.04
AAL	0.99	0.96	0.93	0.95
GA\M	1.00	0.75	1.74	0.98
NWA	1.00	1.04	0.92	0.96
UAL	1.01	1.08	1.04	1.08
Other	1.04	0.76	1.37	0.88
COA	1.05	0.93	1.00	0.95
MEP	1.07	1.11	1.24	1.11
TWA	1.11	1.31	0.99	1.21

There are some interesting points to make about the above results. The first equity metric hovers very close to 1.0 for all carriers. As stated earlier, the differences can most likely be attributed to the carrier's schedule composition.

The third equity metric takes into account the effect of compression. It is interesting to compare the metrics for each carrier without and with compression (equity metric 1 and 3). Some noteworthy points are:

1. "GA/M" went from 1.0 to 1.74 and "Other" went from 1.04 to 1.37. This is easy to explain since these flights are not CDM-participants. The increase for "GA/M"

flights is higher than "Other" because "GA/M" flights are never compressed. "Other" flights are compressed but with a lower priority than CDM participants.

2. There are some carriers, such as AWE, FDX, and MEP, whose equity became worse after applying the effects of compression. This can be explained by 2 different reasons. The first is their cancellation strategies. They did not have as many canceled open slots available in compression. Second, all three of these carriers are not the major carriers into the GDP airports, so their flights are sparse through a GDP. This means that some of the savings attributed to their open slots will be distributed to other carriers, those who acted as a bridge for their open slots.
3. There is one other carrier, UAL, which also has a higher equity value after applying the effects of compression. This is explained by the fact that this carrier is a bridge-only carrier at its seven major hubs. This means that their canceled open slots will not be compressed into, and their compression benefits will be smaller. They prefer to utilize the substitution process instead.
4. The carriers with the biggest improvement in equity due to compression are TWA, USA, and AAL. All three of these carriers are known to use smart cancellation strategies to minimize their initial delays in GDPs. In particular, TWA will wait for compression to fill in their open canceled slots instead of utilizing the substitution process.

These results are also displayed in Figure15. The key equity metric is the first bar of the four for each carrier and hovers very close to 1.0.

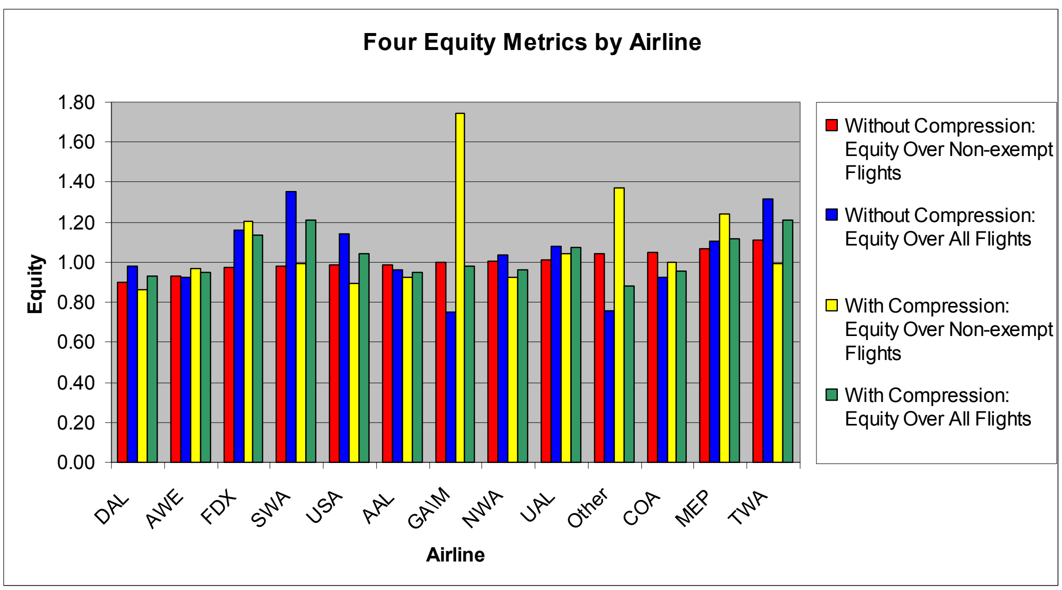

Figure 15. Equity Metrics by Airline

Table 4 displays statistical measurements for each of the four sets of equity metrics.

Table 4. Statistical Measurements for Equity Metrics

	Without Compression:		With Compression:	
Statistic	**Equity Over Non-exempt Flights**	**Equity Over All Flights**	**Equity Over Non-exempt Flights**	**Equity Over All Flights**
Mean	1.004	1.038	1.089	1.031
Standard Error	0.016	0.051	0.069	0.030
Median	1.001	1.036	0.993	0.984
Standard Deviation	0.056	0.182	0.248	0.109
Sample Variance	0.003	0.033	0.061	0.012
Kurtosis	0.415	-0.336	3.307	-1.015
Skewness	0.044	0.106	1.787	0.535
Range	0.212	0.599	0.884	0.330
Minimum	0.900	0.752	0.860	0.881
Maximum	1.112	1.351	1.744	1.211
Confidence Level (95.0%)	0.034	0.110	0.150	0.066

As is shown, the mean, 1.004, and the median, 1.001, for the first equity metric is very close to 1.0, more so than the other three metrics. Also the range, the difference between the maximum and minimum values, .21, is the smallest. This compares with .6, .9, and .3 respectively for the other 3 metrics. The variance, .003, and standard deviation, .056, is also very small.

The conclusion is that based on the equity metrics established here, the RBS algorithm has proven to be a fair and equitable mechanism for assigning arrival slots to flights during a GDP.

3.3 Frequency of GDP-E Revisions

The capability to "revise" a program was not available prior to CDM. This is a new feature under CDM that allows specialists at the Command Center to modify crucial GDP parameters such as the airport acceptance rate (AAR), the scope of the program (centers) or the duration of a program. Prior to the institution of CDM procedures, there were other less flexible mechanisms used. One such mechanism is called a blanket delay. During blanket delays, all flights in a program are assigned equal amounts of delay. This type of action did not allow for the consideration of special issues and were based on a "best guess." The other tool is known as an extension in which a program is just extended in time with all other GDP parameters remaining the same. According to Forrest Terral, changes to programs prior to CDM in many cases resulted in the loss of program integrity.

It is worth documenting the use of this revision tool that allows more flexibility to modify GDPs based on updated forecasts while maintaining program integrity. A decision to revise is always based on better information and this allows for more effective and efficient ground delay programs. Table 5, presents the extensive use of the revision tool.

Table 5. Number of GDP Revisions

Airport	Single Revisions		Double Revisions		Triple Revisions	
	9/97-8/98	9/98-8/99	9/97-8/98	9/98-8/99	9/97-8/98	9/98-8/99
ATL	0	11	0	4	0	3
BOS	0	23	0	4	0	5
EWR	11	25	3	13	2	5
ORD	0	21	0	8	0	13
SFO	11	39	5	27	4	8
STL	6	15	0	5	1	1

This tool is being utilized more frequently during the *CDM* phase. At least 10 log entries in the ATCSCC GDP critique attest to the effectiveness of revisions to smooth out the traffic [and reduce] departure delays. The flexibility of this tool has resulted in the avoidance of underutilized capacity and excessive airborne holding. Since there is no tool comparable to a revision prior to CDM, a quantification of a specific impact of CDM is difficult to ascertain.

Table 6. Average Early, On-time, and Late Departure Percentages

Periods	Average Percentage of Early Departures	Average Percentage of On-time Departures	Average Percentage of Late Departures
Pre-CDM	27.64	50.85	21.51
CDM	17.30	65.87	16.83

The average on-time departure percentage has improved from 50.85 to 65.87 percent. This means that 15.02 percent more flights maintain departure compliance since the CDM period started. This implies a 31.74 percent[11] improvement for on-time departure over the pre-CDM period. This improvement came from the decrease in early and late departures. The average early and late departure percentage has decreased by 10.34 and 4.68 percent, respectively. Airlines always had an incentive to reduce late departures regardless of CDM status. This is probably the reason we see less improvement in this category compared to the improvement in early departure. Nevertheless, the 21.76 percent[12] late departure improvement over the pre-CDM period is an achievement. The improvement in early departures over the pre-CDM period is 37.41 percent[13]. We believe that active information exchange between the FAA and airlines along with paying close attention to the operation accomplished this outstanding improvement.

[11] (Pre-CDM average percentage of on-time departure) * (1.0 + on-time departure rate change) = CDM average percentage of on-line departure; 50.85 * (1.0 + on-time departure rate change) = 65.87; On-time departure rate change = 31.74; Positive rate change indicates the improvement.
[12] (Pre-CDM average percentage of late departure) * (1.0 + late departure rate change) = CDM average percentage of late departure; 21.51 * (1.0 + Late departure rate change) = 16.83; Late departure rate change = -21.76; Negative rate change indicates the improvement.
[13] (Pre-CDM average percentage of early departure) * (1.0 + early departure rate change) = CDM average percentage of early departure; 27.64 * (1.0 + Early departure rate change) = 17.30; Early departure rate change = -37.41; Negative rate change indicates the improvement.

4.0 Status of Initial Collaborative Routing (ICR)

Collaborative Routing can be defined as information sharing for creating and assessing rerouting strategies around hazardous weather, special use airspace (SUA), and other constrained airspace resulting from congestion. Collaborative Routing employs electronic "chalkboards" for use by ATCSCC, en-route Center TMCs, and ATC coordinators at the AOCs for display conferencing. The application of this collaborative "display conferencing" capability results in greater situational awareness, faster decision making, and common understanding of solutions. This in turn means greater flexibility for airspace users in their flight planning. The user may also observe overall flight efficiency, possible resulting in reduction in delay, as all participants have one common view of the constrained airspace improving the quality of their decisions.

ICR functionality continues to be tested in the operational environment. The ICR components: Collaborative Convective Forecast Product (CCFP), Collaborative Routing Coordination Tool (CRCT), and Picture Tel (PicTel), are in different levels of maturity. The following is a summary of the current status as of September 1999, accomplishments, and next steps for each component.

4.1 Collaborative Convective Forecast Product (CCFP)

CCFP is used on a daily basis with positive results. This product has fit into the Air Traffic Control Command Center (ATCSCC) operation with ease and provides an effective collaborative weather forecasting vehicle. Data collected over the last three months supports the effectiveness of the product in terms of accuracy, readiness, and operational impact. The overall accuracy through July 31 is 89 percent; readiness is 100 percent, and operational impact is positive 88 percent of the time.

Over the last few months a new thunderstorm high coverage area of 74 to 100 percent was added to provide a more comprehensive picture of forecast thunderstorm activity. In addition, the geographical coverage area was extended to a point West of Denver, Colorado.

The CCFP process terminated for the year on August 31, 1999, and will resume when the convective activity begins again in 2000.

4.2 Collaborative Routing Coordination Tool (CRCT)

The ATCSCC evaluation of CRCT functions continued for the remainder of the severe weather season. CRCT completed two essential items in its development for FY99. They include the completion of:

- The familiarization exercise at Kansas City ARTCC Traffic Management Unit (TMU), and
- A limited evaluation at ATCSCC Severe Weather Area position.

Next steps for CRCT include:

- Continuation of FAA field evaluations at the ATCSCC and two field sites in FY00, and
- FAA will determine plan for fielding capabilities and continuation of research and development based on evaluation results and other factors.

4.3 Picture Tel (PICTEL)

The new FAA data bridge has been installed and underwent minor adjustments while successfully supporting PicTel conferencing during the week of August 23,1999. Regular use of PicTel commenced on August 30.

The PicTel bridge service has been supporting PicTel SWAP data conferencing. The information exchange, utilizing the CCFP and TSD, has been generally good. The following FAA facilities and Airline Operations Centers are currently participating: ATCSCC Severe Weather Position, Washington ARTCC, New York ARTCC, New York TRACON, Boston ARTCC, Indianapolis ARTCC, Cleveland ARTCC, Delta Airlines, and Northwest Airlines.

User comments to this point have been favorable. The general consensus is that PicTel holds great potential; however, there needs to be a greater focus on content. The intrinsic benefit is that although CCFP and other products are available on various web pages, having the PicTel conference forces the participants to use the same data as part of the collaborative planning process.

The ATCSCC has mandated the use of PicTel data conferencing daily at 1550z from August 30, 1999 through October 15, 1999. Additional “NAS status” constraint information will be introduced into the PicTel conference. Candidate information includes the OIS summary page, the latest ATCSCC Advisory, and the TSD with its routing features.

5.0 Conclusions

CDM GDPE has been operating for more than one year (in the contiguous US). It has become part of the ATC management system and has been widely supported by both the FAA and the user community. We can reason that the ATC system operates "better" with this improved data quality, timeliness, and slot allocation processes. Like any system, GDP-E can be improved further. Measures like IPE, RCI, and compression with associated equity can help us better understand the value of these improvements.

This analytic team has sought a unified approach for analyzing CDM performance. The most natural candidate, and the one that we have tried to pursue, is to collect performance statistics both before and after program implementation and review them for positive trends. However, this type of pre- versus post- analysis assumes that all else remains constant during the implementation or that, at the very least, results can be screened for complicating factors. But the NAS is a highly dynamic environment that does not lend itself to controlled experiments. For several of the metrics, data was not available in the pre-implementation phase. As a result, this pre- versus post- analysis was pursued only when it made sense to do so.

Initially, we had hoped to adopt simplistic metrics to answer ‘bottom-line’ questions such as: “has throughput improved?,” “have travel times been reduced?,” and “are there fewer delays?” Many of these metrics failed for a variety of technical reasons including incomplete, corrupted, or altogether unavailable data. For other metrics, we found that the further we probed, the less certain we were of the impact that CDM should have. For instance, should the number of ground delay programs increase or decrease as a result of CDM? On the one hand, airline-supplied cancellation data should provide more accurate arrival demand predictions and therefore avoid superfluous ground control actions. (This has been documented on several occasions.) On the other hand, the refinement of GDP procedures and the ability to revise parameter settings has provided the community with a more equitable, more resilient tool. Wider acceptance may have encouraged more frequent use. (Indeed, there have been cases in which the air carriers have requested the use of a GDP.)

It is far easier to assess the programs within a corporate environment. Value can be measured simply by the difference of dollars invested and dollars saved or generated. Unfortunately, we don’t have the same unifying objective (profit) or standard of measurement (dollars) in an aviation setting. True, some assessment can be made of air carrier dollars saved in the form of fuel, man-hours, etc. But fuel consumption is not always a primary concern; an airline may knowingly trade fuel consumption for an earlier arrival or a passenger comfort. We have learned that these types of computations are highly conditional - both situation specific and air carrier dependent.

Many FAA systems have sought to ensure safety *and* improve efficiency within the NAS. Since air traffic safety within the United States has an impressive track record, this is not one of the areas that CDM has sought to directly improve. As far as efficiency is concerned, this is generally taken to mean efficient use of NAS resources on a *global* scale, without reference to individual organizations using those resources. But what's inefficient for one user may profit another. The NAS serves the needs of many groups whose objectives change on a daily basis. This makes it hard to find a uniformly applicable metric. In fact, the very premise of measuring economic savings on behalf of the air carriers defies one of the cornerstones of CDM: that each airline should weigh for itself the economic value of alternative actions.

One uniform measurement of progress that the CDM community has agreed upon is minutes of (FAA) planned ground delay that has been avoided as a result of the Compression algorithm housed in ground delay program procedures. This has become the showcase achievement of CDM activities. As of this writing, the CDM program has logged over 4,000,000 minutes in savings of planned ground delay. Not all of this may transfer to minutes saved in gate-to-gate time; ten minutes of savings at the origin airport could be offset by ten minutes in airborne holding at the destination airport. Nonetheless, the opinion of CDM experts is that the true savings from compression is a significant fraction (e.g., about one-half) of the 4,000,000 minutes.[14]

Weather has proven to be the primary culprit in the disruption of airline operations and the foremost cause of FAA-induced control actions. Ironically, this has also proven to be the prime obstacle for analysis of aviation operations. The RCI that we have developed is the first metric to rigorously factor out weather conditions from evaluation of large-scale, multi-participant performance.

Second only to weather, the greatest impediment to this analysis was the difficulty of modeling what *would have* happened had CDM not been in place. The modeling of something as seemingly simple as the number of aircraft that could have landed under different circumstances requires not only highly sophisticated modeling techniques but also the input of dozens (or even hundreds) of parameters, many of which have to be subjectively fixed. This has, in many cases, caused us to significantly alter or even abandon our first-choice metrics. We have guarded against modifying or distorting metrics to the point that the results become uninterpretable. Whenever possible, we have formulated the simplest, most appropriate metric, given the current field of knowledge, and we have documented the reasons that other metrics were deemed inadequate or inappropriate.

The primary mission of this team has been the development and application of CDM-related metrics. But it is incumbent upon us to place our findings in perspective with CDM activities and accomplishments. What has started out as a grass roots effort has developed into the most mature FFP1 core capability. The true benefits of CDM may lie in the communication and awareness that has been established, particularly between the users of the NAS and its service providers. The working relationships and cooperative efforts that it has formed are essential for the promotion of free flight.

Its strength lies primarily in its ability to surmount cultural behaviors that inhibit communication between entities. CDM has managed to consistently deliver on its short-

[14] Based on the commonly accepted (perhaps dated) industry standard of $25.00 per minute, this is equivalent to $50,000,000 in savings to the airline community since January of 1998.

term objectives while maintaining the pursuit for long-term solutions. Specifically, CDM has:

- Created much needed data sources through the submission of real-time airline operational data;
- Established a communications infrastructure, the CDMnet, for the promotion and dissemination of situational awareness amongst all parties in the NAS;
- Provided equity and predictability to ground delay program procedures;
- Provided an open forum for problem identification and problem solving through the participation and coordination of airlines, industry, and the FAA.

To whatever extent possible, we have tailored the body of CDM metrics to reflect CDM activities and accomplishments. The body of metrics continues to grow and evolve with the CDM program. The technological solutions for improving the NAS have in many cases outgrown our ability to assess their impact. This report is as much a treatise on aviation performance metrics as it is an assessment of CDM program status. But much headway can be made in the development of metrics and analytic methods. For continued analysis of CDM program performance the following metrics should be considered:

- Rate Control Index (RCI)
- Integrated Predictive Error (IPE)
- Compression
- Equity
- Slot usage

Although not investigated in this document, usage of CDM decision support tools such as Flight Schedule Monitor (FSM) can further assist decision-making and operations analysis.

It is vital that aviation performance metrics and solutions to problems within the NAS be further developed in a complementary fashion. These will be applicable not only to the CDM program, but other FFP1 programs as well.

This Page is Intentionally Blank

APPENDIX A. Evolution of CDM Metrics

FFP1 Core Performance Metrics

As noted in their December 1998 FFP1 Core Performance Metrics document, the RTCA identified a list of performance metrics that could be used to measure the impact of Free Flight Phase 1 (FFP1) systems to the National Airspace System (NAS). These metrics were identified for all FFP1 capabilities and are associated with the Air Traffic Services (ATS) Performance Outcomes. Each metric as originally identified by the RTCA was intended to provide one measure of impact for each of the ATS Performance Outcomes. One reason for this association is that these Performance Outcomes provide an appropriate measurement of operational performance of new services while considering the impact to the user community. As stated in the ATS Performance Plan (1997-1999):

> "To ensure that ATS is providing and developing the right type and level of services to its customers, it has generated customer-oriented, outcome-based system performance measures. These measures have been formulated in consultation with the user community and represent the operational outcomes desired by the system users that reflect ATS' overall performance across all services that it provides."

The original core performance metrics were chosen based on the availability of data, knowledge, and direction of impact, and availability of similar metrics in the category. Since no preliminary analysis of the metrics was performed at this stage this process represented a first step to the metrics development process.

GDP-E Core Performance Metrics

The GDP-E Core Performance Metrics identified and reported in December 1998 by the RTCA are identified in Appendix B. Since that time, the CDM Analysis Team has evaluated these metrics as satisfactory measures of NAS performance. Many of the metrics were identified because of their potential to be used in a pre vs post analysis of GDP-E. After additional consideration many were determined inadequate for this type of analysis. As a result, several metrics have been removed, some improved, and still others have not changed. Those metrics that have been changed or removed were done so because of difficulties in interpretability, limited availability of data, complexity of analysis, or limited impact identified from preliminary analyses.

Several of these metrics were removed because data was not available before GDP-E deployment to perform a pre vs. post analysis. The ability to perform a pre-post analysis allows greater confidence that if any changes in the observations are identified it is easier to establish a GDP-E causal relationship. This restriction does not eliminate the possibility of conducting a trend analysis of a metric after GDP-E deployment. In fact, this type of study would be ideal for the System Flexibility metric *control by time of arrival*. The reason being that System Flexibility has a limited number of metrics with which to measure system performance.

A number of metrics were removed because of problems of interpretability. For example, the CDM Analysis Team decided that the metrics, *number of substitutions* and n*umber of cancellations,* were to too complicated to identify whether any change would be a benefit or dis-benefit to the NAS. Similarly, any potential change in the number of cancellations during the study period would result in the same problems. Additionally these metrics suffer from the difficulties associated with normalization of data. Changes in the numbers of cancellations and substitutions would require taking into consideration the type of ground delay program (e.g., All West, 6 West, 1st Tier, etc.), airport demand, time of day, intensity of bad weather (e.g., visibility, precipitation, winds, etc.), and aircraft type, among others.

Several of the metrics were removed because preliminary analysis of empirical data has shown no impact from GDP-E. Metrics eliminated for this reason include *planned versus actual departure and arrival times: average and variance*. A preliminary analysis performed by Metron Inc. has *not* shown a significant change resulting from GDP-E deployment. Additionally, *taxi-out times* have not shown any significant improvement for flights traveling to GDP airports. Additionally, several air traffic specialists stated that any impact to this metric is unlikely and further analysis would not be practical. Consequently, this metric was removed from the System Delay category.

The following sections provide a detailed explanation as to why the original metric was chosen, the methodology proposed to analyze any potential impact, results of preliminary analysis, and the reasoning why the metric was changed or removed from the GDP-E list.

Performance Metrics Removed or Changed: A Detailed Explanation

Number and Duration of GDPs

Although SFO and EWR were the first airports to become full CDM airports in January 1998, the information in Flight Schedule Monitor (FSM) was being utilized at all airports to model different scenarios and subsequently to make decisions about flight plans. Thus, CDM and FSM had an impact on all airports from the initial implementation of CDM. According to one ATCSCC questionnaire " … this STL GDP ran much more smoothly than all … GDPs for STL prior to CDM".

Since full prototype operations began in September 1998 at all airports, the analyses are performed for three time intervals that are referred to as the *Pre-CDM*, *Transition to CDM*, and *CDM* phases. The Pre-CDM phase is from September 1996 through August 1997; The Transition to CDM phase is from September 1997 through August 1998; The CDM phase began in September 1998 and is currently being employed. At this point, analysis has been done up to July 31, 1999. The sources of the data for the analyses are the GDP summaries and logs from Metron.

It was conjectured that the *number of ground delay programs* would decrease as time passed due to availability of better, more up-to-date information from the airlines and the flexibility of the airlines to independently make decisions that would affect the need to set up a GDP. NAS users are given advanced warning through GDP advisories of the possible institution of a GDP. This advanced notice affords the airlines the opportunity to make changes to their schedules such as substitutions, delays or cancellations, which may result in demand reductions. In fact, ATCSCC specialists have stated on several occasions that airlines have reduced demand to such a degree that a GDP was avoided.

There are other factors that affect the number of ground delay programs such as weather patterns in different years. If the weather is more adverse in one year than another, then there may be a need for more GDPs. Additionally, there is evidence that GDPs are not only being implemented in response to inclement weather conditions or in support of severe weather avoidance procedures (SWAP), but also being employed to address other air traffic management problems.

The fluctuation in the numbers of GDPs presented in Table A-1 may be due to varying weather conditions or the aforementioned factors that affect the number of GDPs. Hence, no conclusions can be drawn about the impact of CDM on the number of ground delay programs.

Table A-1. Number of Ground Delay Programs

Airport	9/96-8/97	9/97-8/98	9/98-8/99
ATL	19	15	31
BOS	47	46	56
EWR	71	56	78
ORD	54	47	70
SFO	152	162	180
STL	68	61	42

It is uncertain whether CDM should have an impact on the *duration of ground delay programs*. Results from a previous analysis show that there has been an increase in the absolute differences between the originally planned and actual duration of GDPs prior to CDM prototype operations. This metric is strongly affected by policies at the ATCSCC during the given years along with other factors such as adverse weather conditions. During years prior to and including 1997, specialists could not plan programs lasting longer than 4 hours and during years after and including 1998, they could not plan programs lasting longer than 6 hours. Since this metric is affected by factors other than CDM, it was decided that this metric is ambiguous.

Percentage of GDPs Canceled Near Start

It was hypothesized originally that any cancellation of a ground delay program is an indication of an inappropriately planned program, and therefore, the percentage of canceled programs should decrease as time passed through the different phases. Figure A-1 illustrates that there is erratic behavior during the *Transition to CDM Phase* (9/97-8/98), but comparison of *Pre-CDM* to *CDM* indicates that there has been a decrease in the percentage of canceled GDPs. This metric is again strongly affected by weather conditions and forecasted demand. Cancellations of GDPs occur as soon as it has been decided that there is no need for a GDP, which is sometimes based on better weather and demand information. Since there are instances when a cancellation of a GDP is a desirable action, further stratification of cancellations is needed to extract those that are indeed undesirable. Therefore, this metric has been replaced by *the percentage of GDP cancellations near start*.

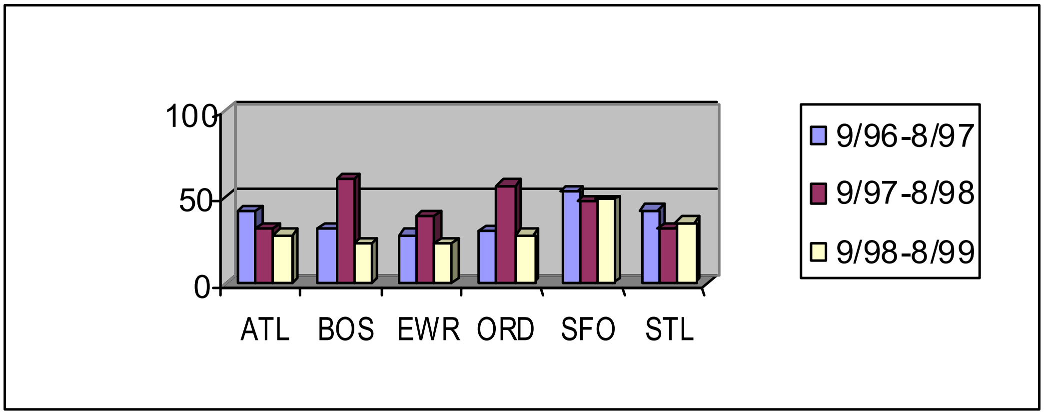

Figure A-1. Percentage of Cancelled GDPs

Average Taxi-out Time

The metric, *mean taxi-out time,* was identified by the RTCA working group because of the possibility of GDP-Es favorably impacting the number and duration of aircraft being held on airport runways. Specifically, aircraft on the tarmac awaiting departure to airports that are under a ground delay program might benefit from GDP-E by reducing taxi-out time. It was thought that the underlying cause of this benefit was that because of better and more timely information provided by the CDMnet, some aircraft might avoid being caught on the tarmac during a ground hold – reducing taxi-out times. Of course, even if this theory is valid, overall delay time of aircraft would likely not change. Taxi-out delay would simply shift downstream to the gate. The net effect would likely be zero.

A statistical comparison was made pre- and post-GDP-E looking at mean and variance of taxi-out times. This preliminary analysis looked at aircraft departing from John F. Kennedy Intl. (JFL), St. Louis Intl. (STL), Las Angeles Intl. (LAX), and Chicago Intl. (ORD) during an SFO ground delay program. The results of this analysis were briefed to the CDM subgroup for their feedback. In general the group felt that the theory behind the hypothesis that taxi-out times would decrease with GDP was optimistic but unlikely. The purpose of GDP-E was never to reduce taxi-out times and even if on rare occasions when GDP may prevent an aircraft from beginning its taxi-out when it would have without GDP-E it is unlikely that enough of these instances would occur to should changes in mean and variance with such large data sources being analyzed.

Additional problems were noted by the CDM subgroup including the necessity to account for the various types of ground delay programs. (That is, those GDPs that effect all centers, nine west centers, six west centers, no west centers, etc.) Furthermore, it was stated that the analysis would have to be normalized for type of weather conditions and possibly even demand. As a result of all of these challenges, taxi-out time was removed from the CDM subgroup's metric list.

Flying Time, Distance, and Fuel Usage

Since the control method of a GDP is in the time domain, the metric *flying time* was identified and used to measure the operational impact of GDP-Es. It is believed that *flying time* is appropriate for measurement, since it is a good representation of two higher-level categories, Delay and Predictability. The metric *flying distance* was discarded since it is believed to be a more appropriate metric for measuring the impact of Initial Collaborative Routing (ICR). The metric *fuel usage* is retained, but used simply as a multiplier (gallons or dollars) applied to flying time.

Flying time is defined as a measure of holding or delay absorption associated with arrival runway usage. Using two circles with specified distances from the arrival airport, the traversal time from the outer circle to the inner circle is assessed. From examination of operational data and discussions with knowledgeable personnel, rings at distances of 100 and 40 nmi are appropriate. That traversal time is assessed against a baseline. This baseline is defined as the ideal, modeled traversal time based on the flight trajectory software kernal in the Collaborative Routing and Coordination Tool (CRCT). Also assessed is a VMC baseline, to evaluate the holding or delay absorption which is the standard air traffic management practice in a certain locality.

The excess arrival flying time as defined above can be interpreted as follows. The number of flights subjected to excess arrival flying time, and the amount of the excess might increase under a GDP because of uncertainty in the true airport landing capacity, and the reduced controllability of take-off times. To the degree that the GDP excess arrival flying time is like the VMC value, then the GDP is successful, since it keeps pressure on the runway via appropriate airborne reservoir. Conceptually, we want to ensure that holding does not increase as the ATM system seeks to fully utilize scarce runway capacity.

Two sources of data were used: Enhanced Traffic Management System (ETMS) data (for flight time and position information), and ATCSCC program logs (to know the start time, duration, etc. of the GDP).

The metric is appropriate for two uses: either to (1) assess the performance of GDP-E, relative to GDP, prior to enhancements or, (2) ongoing monitoring of GDP-E. This section generally examines use (1). Regarding terminology in this section, "GDP" refers in the generic, to a ground delay program, the traffic flow management initiative used by the FAA to limit the number of arrivals during reduced capacity situations. The terms "GDP-E" and "post-GDP-E" are used synonymously, meaning the "Enhanced" Ground Delay Program (i.e., including collaboration and data-sharing). The term pre-GDP-E refers to the simpler method, commonly called "Grover Jack."

Many factors influence this metric: weather conditions, air traffic demand level (and direction of flow), operational personnel, aircraft type mix. These are handled as follows. For aircraft type mix, we have removed lower-performance aircraft from consideration. These have a higher variability in speed, and can fly more slowly without actually being in a delay situation. The metric should be considered with this caveat, that not all flights are included in the metric. Appendix D shows the equipment types, using ETMS codes, which, as a group, generally have filed speeds high enough that a slow-down can be detected using the method employed here.

Regarding direction of arrival flow, initial tests for various airport arrivals indicate that arrival flows along the differing Standard Terminal Arrival Routes (STARs) exhibit very similar range of traversal time. The other influences - weather, demand level, and personnel - should be considered when comparing GDP-E to pre-GDP-E. For best comparison, a matching of these influences should be undertaken, so that like conditions are present in contrasted scenarios.

We suspect that even with our attempt to match days for comparison, the system is replete with extraneous noise, which would obscure a significant difference, if one indeed exists. Just considering the "physics" of the problem, if the number of flights showing up for arrival services exceeds the safe capacity of a facility (approach control or tower), some type of flow control or delay maneuver *must* be undertaken. But given the complexity of the NAS, and the additional dynamics and complications brought on by a GDP, we are faced with the problem of interpretability of our data. Given the data analyzed here, we cannot conclude that delay maneuvers for airborne flights are either lower or higher under GDP-E.

Number of Cancellations and Substitutions

The *number of substitutions* was originally chosen as a GDP-E performance metric. Substitution is described as the exchange of arrival slots for certain flights. Simplified substitutions, is described by the CDM Working Group as

> …the need to identify specific pairs of exchanges or substitutions (e.g., flight one is canceled and flight two is substituted into flight one's arrival slot) is eliminated. Users will be allocated a set of arrival slots, and in the initial solution, there will be an initial assignment of flights to slots. If a user cancels or delays a flight that would change the slot assignments, the user simply will report that flight two is now assigned to slot one, flight three is assigned to slot two and so forth. The capability to conduct simplified substitutions is being embedded in the CDM message structure.

This GDP-E metrics presents the problem of interpretation. It is not known if the change in number of substitutions is good if it increases or decreases. The CDM Working Group is currently investigating the impact of any change in the number of substitutions and will report on the results of the investigation once completed. However, since any potential change in the number of substitutions cannot be read as either positive or negative, this performance metric has been removed from the primary list.

The metric, *number of cancellations,* suffers from the same problem of interpretability as the number of substitutions. If the number of cancellations change it is not known whether an increase or decrease is a good thing. Some CDM analysts believe that cancellations to GDP airports will fall as a result of GDP-E because of the ability of the airlines to better employ arrival slots. Others believe that the number of cancellations may go up or even stay the same but the characteristics of the cancelled flights will be different. That is, if airlines have more time to determine which flights they want to cancel they may cancel the low revenue flights more often and reallocate the high revenue flight to guarantee an arrival slot and even a reduction in delay. Without the help of the airlines it is uncertain whether the evaluation team will be able to determine individual characteristics of certain flights. As a result, the numbers of cancellations has been removed as a primary metric.

APPENDIX B. Evolution of GDP-E Core Performance Metrics

Operation Outcome	Former Metric	Status/Change in Metric	Current Metric
System Safety	*Number of operational errors*	No change.	***Number of operational errors*** [b]
	none	Number of operational deviations (as defined in *Air Traffic Quality Assurance (7210.56A).* Chapter 5, Section 1) was added as a System Safety metric.	***Number of operational deviations*** [b]
System Delay	*Average flying time*	No change.	***Average flight time*** [ab]
	Number of GDPs	Removed as a delay metric and reintroduced as the number of GDPs canceled under System Predictability.	***observation*** [d]
	Duration of GDPs	This delay metric was removed because of the difficulty in normalizing for all possible GDP-type factors.	***observation*** [d]
	Average difference between planned time versus actual time (arrival, departure time)	Removed from System Delay category. Preliminary analysis of empirical data has shown that there is no impact to this metric.	***Rate Control Index (RCI)*** [b]
	Average taxi-out time	Removed from System Delay category since ATCSCC representatives believe GDP-E impact to taxi-out time is unlikely.	***observation*** [d]
	none	Compression minutes saved, the cumulative reduction to scheduled delay- not actual delay, has been added as a System Delay metric.	***Compression minutes saved*** [a]
User Access	*Number of operations*	No change.	***Number of operations*** [b]
	Number of unused slots	This metric was removed since the number of unused slots was not collected prior to GDP-E. As a result, a pre-post comparison cannot be performed.	***observation*** [d]
	Number of cancellations	This metric was removed since any potential change in the number of substitutions cannot be defined as either good or bad.	***observation*** [d]
	Number of substitutions	This metric was removed since any potential change in the number of substitutions cannot be defined as either good or bad.	***observation*** [d]

System Flexibility	*Average flying distance*	No change.	***Average flight distance*** [c]
	none	Control by CTAs was added as a metric for System Flexibility. Given an arrival slot, the user may determine its own departure time according to its own economic objectives. Since it is a new CDM functionality it will not permit a pre-post comparison but given the limited number of metrics in this category it will be studied.	***Control Time of Arrival*** [b]
System Predictability	*Variability in planned versus actual arrival/departure time*	No change.	***Variability in planned versus actual arrival/departure time*** [a]
	Variability between arrival rates (planned and actual)	Measures how close the flow of traffic into an airport matched the targeted pattern of traffic, both in quantity and distribution.	***Rate Control Index (RCI)*** [b]
	Average fuel usage	Removed as a predictability metric because of the difficulty of normalizing for all GDP-type factors.	***none*** [c]
	none	This metric assigns a single number to the stream of predictions based on how far off (in absolute value) the predictions were from the actual event.	***Integrated Predictive Error (IPE)*** [a]
	none	EDCT compliance ratio can measure the impact of better and more timely information to airlines possibly contributing to greater EDCT compliance and a more predictable NAS.	***EDCT compliance ratio*** [a]
	none	Since the number of GDPs is highly dependent upon weather, the number of GDPs canceled and revised was chosen as possible measures of GDP-E impact to ground delay programs.	***Number of GDPs cancelled near start, number of GDP revisions*** [a]
	none	This metric was added in an attempt to measure excess flight distance in the terminal environment, possibly due to excess time in a holding pattern.	***Variability in distance flown*** [b]

[a] Metrics that have been evaluated in this report.
[b] Metrics that may be investigated in future studies.
[c] Metrics that under analysis have been found to be inadequate as performance metrics.
[d] Metrics that may be valid but difficult to interpret. They will nevertheless be observed during the course of FFP1.

APPENDIX C. Air Carrier Responses to the Questionnaire

MEP – Midwest Express Airlines

Q: What positions at your carrier are utilizing FSM?
A: Operations Coordinator and to some extent the dispatchers. They use it to look for possible holding.

Q: What positions are utilizing AADC (Airport Arrival Demand Chart)?
A: Not sure what this is

Q: What is FSM being used for?
A: To monitor and assess the impact of a ground delay program on the airline.

Q: What is AADC being used for?
A: Right now, we do not use AADC regularly. We had experienced numerous situations where the data was old and are not sure we can trust it.

Q: How has access to CDM information changed your operation?
A: It gives a real-time view of what's going in the NAS. It allows us to tailor our operation as conditions change.

Q: How had access to CDM information changed your relationship to TFM / ATCSCC?
A: By allowing us to have the same "sheet of music", we are able to interact a lot better with ATCSCC

Q: Have you realized any benefits from CDM that you did not originally anticipate?
A: NO, all benefits realized were originally anticipated -- this is probably due to our lengthy participation before becoming a member... we had a good idea as to what to expect.

Q: Do you have any "success stories" of particular GDPs where CDM provided a benefit to your operation?
A: Not sure.

Q: Have you done any analysis to estimate dollar savings or other benefits due to CDM?
A: We have estimated for the first 2 1/2 months we saved approx. $50,000 + due to reduced delays because of compression. This figure was derived using a $25/delay minute times the number of minutes our gateholds were reduced by compression.

Q: Do you have any general comments on CDM and how it has effected TFM or your carrier?
A: The sharing of information has always been a good concept. It leads to a better, quicker and more efficient decision making. Keep the information coming......

GAA – Business Express Airlines

Q: What positions at your carrier are utilizing FSM?
A: The Operations Coordinators, who are the "Duty Managers" and are in charge of all operational decisions on their shift, primarily use the FSM. The Dispatchers have used the FSM to guesstimate the holding potential into an airport.

Q: What positions are utilizing AADC (Airport Arrival Demand Chart)?
A: I am not familiar with this acronym

Q: What is FSM being used for?
A: The FSM is used to monitor AAR and to predict potential holding, ground stops, and delay programs. We also use the "Generate EDCT List" function to generate a list of data that is export/import to another program for use in doing slot substitutions (we "grab" the list and import it into Excel). We will generate a new list several times each hour, so the most current information is in our Sub program. Because we use this EDCT list, it is very important that the list be updated at all times. We find that the list is not always updated when a program is put out, and we have to call the ATCSCC to request an update.

Q: What is AADC being used for?
A: Again, I am not familiar with this.

Q: How has access to CDM information changed your operation?
A: The CDM information has allowed us to reduce cancellations and delays on weather days significantly. No real analysis has been done, but the results are noticeable. We are able to plan better and work more proactively. We have also begun doing our own slot substitutions, which has been the biggest contributor to our better performance.

Q: How had access to CDM information changed your relationship to TFM/ATCSCC?
A: We have been communicating with ATCSCC more since we have had access to CDM data. We have noticed that, at some airports, the tower does not allow us to taxi or get clearance in time to meet the EDCT (wheels-up!). Other times, a tower does not get a new (subbed-up) EDCT updates, so we have to call ATCSCC and request that they contact the tower and release our flight. This takes valuable time and resources from our operation and from ATCSCC. Also, there continues to be problems when towers do not release our flights from ground stops, when a GS cancellation message is sent out.

Q: Have you realized any benefits from CDM that you did not originally anticipate?
A: No real analysis has been done, but we did expect some improvement after CDM, and have had some improvement.

Q: Do you have any "success stories" of particular GDPs where CDM provided a benefit to your operation?

A: There have been GDP's (no specifics dates) where we were able to sub up our flights to a point where we could actually manage the delays in the GDP, and after a compression, have noticed even more reductions (bonus!).

Q: Have you done any analysis to estimate dollar savings or other benefits due to CDM?
Answer: No

Q: Do you have any general comments on CDM and how it has effected TFM or your carrier?
A: The CDM process has improved our performance during GDP's, and seems to make GDP's less detrimental to our operation. We notice that GDP's are not planned for a long enough periods of time, even when everyone knows that the program will be extended. We are not able to sub flights after the program's end time, so when the program is extended, the RBS++ compresses our slots before they can be used to sub our priority flights. Extensions to program hit us harder than the original program, because we don't have the option to cancel more flights and sub slots. The RBS++ compression uses our slots from the earlier program. We prefer longer programs (in time frames, not delays!) that include all flights into the night, so we can plan appropriately and sub as required. Compressions/revisions to a longer program are better than extensions to a short program.

TWA – Transworld Airlines

Q: What positions at your carrier are utilizing FSM?
A: FSM is currently at the Traffic Management Unit / Shift Director position and Flight Information Coordinator at STL station. We have also developed an FSM web page so that each dispatcher can monitor his/her flights into STL. They have numerous options they can view: EDCT list/arrival rate/past EDCT/arrival compliance/all arriving flights which also shows AFIX and schedule variation. Web page automatically updates every 5 minutes.

Q: What positions are utilizing AADC (Airport Arrival Demand Chart)?
A: AADC is at the Traffic Management Position.

Q: What is FSM being used for?
A: FSM is being used to:

a. Closely monitor the demand at STL for potential problems such as airborne delays.
b. When airborne delays do occur we use the airborne holding feature to determine the amount of delay.
c. We now are using FSM for planning fuel policy at specific airports.
d. We are now looking at FSM for possibly balancing arrival fixes.
e. When ground delay program advisories are sent we model the program to see what the impact may be before EDCT times are sent. We also do the same with compression which has given us some early heads up on reduced times.

f. We also run extensions to ground delay programs to see how our open slots will impact the extensions of programs. Do we need more canceled flights to reduce delays, etc.
g. When ground stops are implemented we use the ground stop feature and then send out a list of flights captured in ground stops.

Q: What is AADC being used for?
A: We are not really using AADC.

Q: How has access to CDM information changed your operation?
A: Well as you can see from my answers to 1 and 3. The information provided from CDM is being fully utilized at TWA. Compression has especially been helpful at stations like EWR, DFW, ATL, and ORD where we are small players. We are currently getting ready for CR high and low. That is we are loading CR routes in flight planning system and putting publications out for flight crews.

Q: How had access to CDM information changed your relationship to TFM/ATCSCC?
A: When ever I talked to ATCSCC about a particular problem at a specific airport such as ground delay programs/white hats/compression I am looking at FSM.

Q: Have you realized any benefits from CDM that you did not originally anticipate?
A: CDM has of course produced the software FSM which we have become so dependent on that if its not working we feel lost. The ability for the Command Center to quickly revise a program or go from a ground stop to a ground delay program. The CR program is also becoming a reality which I thought would take years and years. The web pages provided by the command center and Metron are also products we never visualized (good stuff).

Q: Do you have any "success stories" of particular GDPs where CDM provided a benefit to your operation?
A: One quick story. We had a program for EWR about a week ago and we had approximately 90min delay. The command center advised they were going to run compression at a specific time. I quickly ran compression on our end and our delay was reduced to down to 20mins. I repeated the drill to make sure I was right and then advised the station to get flight ready to go that the delay would be reduced. When compression was run I missed the time by 4 minutes.

Q: Have you done any analysis to estimate dollar savings or other benefits due to CDM?
A: No.

Q: Do you have any general comments on CDM and how it has affected TFM or your carrier?
A: No.

FDX – Federal Express

Q: What positions at your carrier are utilizing FSM?
A: The ATC Coordinator and Control Center Manager have access to FSM.

Q: What positions are utilizing AADC (Airport Arrival Demand Chart)?
A: AADC is available to all Dispatchers and service recovery specialists.

Q: What is FSM being used for?
A: To monitor the NAS and anticipate Traffic Management Initiatives and their potential impact on our Operations prior to implementation.

Q: What is AADC being used for?
A: Excellent tool for planning holding fuel.

Q: How has access to CDM information changed your operation?
A: Allows us to make better informed decisions. Also allows us to prioritize flights in a dynamic manner according to our business needs at that point in time.

Q: How had access to CDM information changed your relationship to TFM/ATCSCC?
A: A better understanding of why programs are enacted. Also cuts down on landline inquiries.

Q: Have you realized any benefits from CDM that you did not originally anticipate?
A: Not yet.

Q: Do you have any "success stories" of particular GDPs where CDM provided a benefit to your operation?
A: On a couple of occasions through selective swapping we have been able to completely mitigate delays in our priority one positioning flights.

Q: Have you done any analysis to estimate dollar savings or other benefits due to CDM?
A: Not Yet.

Q: Do you have any general comments on CDM and how it has affected TFM or your carrier?
A: CDM is going to prove invaluable as far as providing decision makers in AOC's common situational awareness. I challenge anyone to show me a program that has brought more quantifiable results to the table. Its time to remove the shackles and let this group work on additional initiatives outside of the FFP1 guidelines.

In addition to the survey, Bill Leber from Northwest Airlines sent an email about a snow event in MSP. His email states “The 8th largest snow event ever in MSP happened March 8 and continued into March 9th. On the 8th FSM allowed us to delay the onset of a GDP until MSP went below minimums. Ground stops ensued and then a GDP was issued with a lower rate to allow release and recovery of diversions. Coordination with ATCSCC, the MSP ATCT and TRACON as well as the MSP airport authority was intense. The use of FSM was continuous and vital to virtually every discussion including discussions of timing snow removal activity to fit slack demand. A major benefit was gained when the program was compressed at 0022Z. We saw some flights save hours of delay as cancellations continued throughout the event. The largest benefit was not however the delay of the program nor the compression, not even the fact that the next day we used FSM info to negotiate the cancellation of a MSP GDP (issued at 1430Z, canceled at 1500Z based on NWA thinning of scheduled flights) saving 1100 minutes of delay. The biggest benefit came on March 9th when we were in position to operate a fairly normal schedule. In other words we were able to save tomorrow's operation which is what airlines are forever trying to do during irregular operations. I have asked my Ops Analysis people to attempt to quantify the value of a second order benefit when the disruptions of the previous day are compartmentalized an do not spill over into the next day. It is unfortunately a highly subjective question for analysis."

Command Center Results

The following results contain direct quotes taken from the ATCSCC's logs. The date and airport in which the GDP was run is referenced.

Date	Airport	Comment
6/25/98	SFO	The program delivered steady traffic without significant bunching. According to UAL, the program worked well and ran pretty smoothly. All documentation and coordination timely and complete.
6/26/98	SFO	Overall the program ran well and kept pressure on the airport.
7/1/98	SFO	The program delivered the appropriate rate and ran to completion.
7/23/98	SFO	The multiple compressions run today worked fine with no adverse items noted. It is important to also note there was no "rolling spike" of arrivals or arrival demand above the AAR that would cause us not to compress.
8/9/98	LGA	The hourly arrival numbers were consistent with the program objective.
8/16/98	LGA	DCC modeled many GDP scenarios on FSM. The compression and revisions were timely and the users were kept well informed of the changes that were made. LGA reported that things went real well. Program run very smoothly.

8/23/98	LGA	The program was not compressed because of airborne holding delays. The program was revised, using the same AAR, to make up for the ZDC and ABW ground stops. The big problem this evening was the 'rolling spike'. This spike makes it hard to make a sound decision on compression/revisions. In addition, when we revised the program with the same AAR, the max delays went from 64 to 121 and the Ave went from 33 to 79. The landing numbers were good, most were in the ball park. We had two hours when LGA landed less than the program AAR but during this time frame we were running TEB's and holding LGA traffic. ZNY, ABW, and ZDC held on and off all night. This holding was not caused by the program AAR, but was due to the conflicting approaches at LGA and TEB. LGA was real happy with the way things went. Good demand for both arrivals and departures all night long. N90 said it was a good night. The program treated them well. ZNY reported that the program worked well for them. They held on and off all night but
9/22/98	SFO	Program worked well, desired numbers delivered, FSM worked well.
10/4/98	LGA	Program ran very well. Landing numbers were consistent with the program rate. The decision was made not to compress during the first half of the program due to demand already exceeding capacity. This was evidenced by the FSM numbers and minor airborne holding. Additionally, any unfilled arrival slots would be used by departures (single runway operation.)
11/2/98	STL	ATCSCC supervisor commended that this STL GDP ran much more smoothly than all day GDPs for STL prior to CDM.
11/30/98	SFO	"The program ran well, SFO arrival numbers were well in line with the AAR and program rate."
12/13/98	ATL	"The program delivered appropriate numbers once it was established and ran smoothly… No negative facility or user complaints were received."
12/20/98	STL	"The implemented GDP ran extremely well… User comments were positive."
12/21/98	STL	"Program was run with a 32 rate and a GA factor of 2. STL reported hourly arrival counts, fairly consistently, of 30, though several compressions were performed."
12/24/98	LGA	"The program worked very well, keeping pressure on the airport without going into holding and without incurring departure delays. The landing numbers were in line with the program goal."
1/7/99	ATL	The GDP critique states "The program worked well in smoothing out the demand and delivered the appropriate flights for the rate."
1/7/99	SFO	The 2nd GDP critique in the ATCSCC log states "The

		program initially delivered light and was compressed at 2229Z. Average delays reduced 25 minutes. Program ran well following the compression."
1/9/99	ATL	The ATCSCC GDP critique states: "Program ran well, keeping demand on airport."
1/17/99	STL	The ATCSCC GDP critique states: "The program ran a little light in the 2000Z and 2100Z hours following the revision. Two compression of the program helped increase the arrival pressure on the airport."
1/19/99	SFO	The ATCSCC GDP critique states: "With the exception of the 0000Z hour (the time of the program revision), the arrival numbers met the program objective. ZOA said that it ran as expected but the weather did not fluctuate as they had hoped to allow for a 45 AAR and that was the reason that the program was revised back to a 27 AAR."
1/21/99	ORD	The ATCSCC GDP critique states: "The program delivered in line with the program parameters and was canceled at 0230 based on improved weather and increased AAR."
1/22/99	STL	The ATCSCC GDP critique states: "STL actual arrival numbers were in line with the program rate."
1/25/99	EWR	The ATCSCC GDP critique states: "The GDP was compressed once, and delivered appropriate numbers."
1/28/99	EWR	The ATCSCC GDP critique states: "This program ran remarkably well. Compressions were performed when warranted. Landing numbers were in line with the program rate. Arrival delays were minimal."
1/31/99	ATL	The ATCSCC GDP critique states: "Overall the program with the revisions and compressions worked as well as conditions dictated."
2/7/99	PHL	The ATCSCC GDP critique states: "The TMS monitored the program closely and compressions were done in a timely manner to keep pressure on the airport. The program did run light for one hour, but overall the GDP ran very well."
2/18/99	EWR	The ATCSCC GDP critique states: "GDP well designed and aggressively managed to make the most of a changing situation."
2/20/99	SFO	After the early revision and compressions, the program ran well. The ATCSCC log states at 0110 "Talked to ZOA, GDP running well, no holding, but good pressure on the airport."
2/25/99	BOS	The ATCSCC program digests states at 2012 "We decided to cancel the NGDP and use MIT/ground stops when necessary. The FSM data indicates well over 300 flight cancellations, after modeling a release of the EDCTs. The FSM shows mid/high twenties for demand. A conference with the airlines confirms the cancellations and also indicates there are some flight cancellations that don't even show up. This info helped

		us make the decision for the NGDP."
2/28/99	BOS	Finally, the ATCSCC GDP critique states: "The BOS program turned out to be the most work intensive GDP all night. [due to] The changing conditions at the airport …The program work[ed] as intended by delivering the appropriate number of aircraft as per the rate."
2/28/99	LGA	The ATCSCC GDP critique states: "Overall the program ran well as appropriate arrivals were delivered as per the program rate."
2/28/99	PHL	The ATCSCC GDP critique states: "This program ran well. There was a period in the 2100 hour where PHL was able to use two runways and they landed 46…USA was pleased with the operation."
3/4/99	BOS	The program digest states at 0059: "All comments from the ZBW/BOS TWA was that the program worked very well."
3/6/99	EWR	The ATCSCC GDP critique states: "All facilities/users indicate program is working well, airborne holding is down and users are able to sub."
3/7/99	BOS	The ATCSCC GDP critique states: "This program ran extremely well. The landing rate was slightly above the program rate for all except the 2100Z hour in which runway 33L was closed for 15-20 minutes. Airborne holding kept pressure on the airport without adversely impacting ZBW. No ground stops or revisions were needed."
3/14/99	SFO	Finally, the ATCSCC GDP critique states: "The landing numbers were consistent with the program rate. Arrival delays were within the acceptable range. ZOA was very happy with the way the program delivered the aircraft."
3/16/99	SFO	The ATCSCC GDP critique states: "32 rate and compressions combined to keep a good, adequate demand on the airport. Numerous periods of holding but only one period went in excess of 24 minutes. GDP well managed throughout. Good coordination with major users."
3/18/99	LGA	The ATCSCC GDP critique states: "The GDP was extended and delivered appropriate numbers."
3/19/99	SFO	The critique further states: "GDP appears to have performed well, arrival numbers consistent with rate, minimal arrival holding."
3/25/99	SFO	The ATCSCC GDP critique states: "GDP delivered aircraft as expected."
4/1/99	EWR	The ATCSCC GDP critique states: "GDP delivered expected numbers."
4/1/99	LGA	The ATCSCC GDP critique states: "GDP worked well in delivering program goals."
4/1/99	PHL	The ATCSCC GDP critique states: "GDP delivered appropriate numbers and was canceled as soon as practical."

4/8/99	SFO	The ATCSCC GDP summary states at 1800: "Program continues to run smooth, demand is equal to the developed amount."
4/9/99	EWR	The ATCSCC GDP critique states: "Program did what it was designed to do, and that is to support SWAP, which it did."
4/11/99	CVG	The ATCSCC GDP digest states at 0115: "Shift summary: . . .No arrival delays or holding were reported. All went well, and this was probably the smoothest operation in the country tonight. Facility comments: CVG was very happy and pleased with how the program ran, since they did not have to fight about holding and arrival rates and their operation went smoothly. ZID said it was a very busy shift with all the programs. Ron stated that the CVG program was wonderful."
4/16/99	CVG	The ATCSCC GDP critique states: "The program was compressed throughout the evening with the program delivering in accordance with established rates."
4/16/99	PHL	The ATCSCC log states at 0009: "USA (Bruce) ... states the GDP worked perfectly to manage the flow of USA flights into PHL ..." The ATCSCC log states at 0130 "PHL (Steve): Holding is expected when thunderstorms arrive and the holding went ok, no holding while GDP in place, from the TMU perspective, things went well."
4/28/99	ATL	The ATCSCC GDP critique states: "The program worked well to deliver appropriate numbers of aircraft."
5/4/99	ORD	The ATCSCC GDP critique states "The program ran well and delivered at or near the program goal during all but the second hour. This low hour could be a result of the severe weather reroutes that were in effect."
5/7/99	EWR	The GDP digest states at 0240 "The EWR program delivered accurately."
5/7/99	LGA	The ATCSCC GDP critique states: "The program delivered the appropriate number of arrivals and only one compression was implemented . . . Although the program was implemented for conditions at the airport, it worked well in supporting enroute severe weather conditions by helping in controlling volume."
5/11/99	ORD	An extension of the GDP was avoided due to the number of cancellations displayed in FSM. The ATCSCC GDP critique states: "Discussed the outlook for this evening at ORD with ZAU and the airlines. Arrival volume has been reduced after 2259Z due to the number of flight cancellations...After reviewing the remaining demand and the ORD IFR arrival counts thus far in the program, the decision as made to not extend that program and to let it expire at 2259Z."
5/11/99	ORD	The ATCSCC GDP critique states: "The program ran to completion at 2259 and consistently delivered arrival volume at or slightly above the airport acceptance rate of 80."

5/18/99	EWR	The GDP critique states: "The program ran well and assisted in keeping the flow manageable in support of severe weather."
5/18/99	LGA	The ATCSCC GDP critique states: "The program was revised prior to release [of the ground stop] and worked well to deliver the program goal. The GDP also assisted in keeping the flow manageable in support of severe weather."
5/28/99	SFO	ATCSCC log states: "Facility Check: FSM program running great. Continuous pressure on the airport without holding. Good program."
6/8/99	ORD	The ATCSCC GDP critique states at 0145: "Program continues to deliver within planned parameters. 0000Z hour 65 arrivals, 0100Z hour appears to be in the low 70's, airborne holding under +15."
6/11/99	ATL	The ATCSCC GDP critique states: "The program was managed and delivered consistent with established goals."
6/13/99	BOS	The ATCSCC GDP critique states: " During the program time frame the appropriate number of arrivals were delivered as per the rate."
6/19/99	LAX	The GDP digest states at 1930: "Conferenced ZLA/SCT for program feedback. Both couldn't be more pleased. For several days they have experienced excessive delays along with no equity. Today's program definitely put a handle on the traffic situation. The demand of 91 in the 1700Z hour definitely needed to be addressed. It would appear the program will accomplish what was intended, to keep departure delays equitable and not to have a backlog of departures in the ZLA 1st tiers, or experience excessive departure delays."
6/21/99	EWR	The ATCSCC GDP summary states: "Program delivered appropriate numbers and ran to completion. ... Users and facilities gave positive feedback on the GDP."
6/21/99	EWR	At 00020 the ATCSCC log states: "COA ... wants to know if there will be any holding. Not anticipating any and if there is, it would be a surprise because the GDP has been running so well. Diane [COA] added it ran very good. No complaints from COA."
6/21/99	EWR	At 0025 the ATCSCC log states: "N90/EWR ... just wanted DCC to know how well the they thought the GDP ran. They had too many aircraft, we ran a GS into a program - perfect."
6/22/99	ORD	The ATCSCC log states at 2305: "Facility Check: ZAU/ORD - The revised program seemed to smooth out the traffic, departure delays are diminishing and things are running fine."
6/23/99	STL	ATCSCC log states: "STL felt it [the program] was good and there were no problems with the amount of aircraft that were delivered."

6/27/99	LGA	Regarding program performance, the ATCSCC GDP critique states: “The AAR was set at 27 and the actual arrival numbers came very close to the DCC goal. The program kept holding to a minimum and kept pressure on the airport. It also reduced the possibility of no notice holding which kept ZOB out of trouble and avoided backups to western centers. ZNY reported the program ran very well. It reduced the volume which helped them to deal with deviations. Overall the program did what it was designed to do and was handled very well by our specialists.”
7/2/99	BOS	The ATCSCC log states: “The GDP worked well to exit the extensive Ground Stop in an orderly fashion. The GDP eliminated the need for MIT and the possibility of further Ground Stops due to excess airborne volume.”
7/6/99	EWR	The ATCSCC log states: “GDP appropriately coordinated, calculated and implemented for severe weather avoidance procedures support.”
7/6/99	LGA	The ATCSCC log states: “GDP appropriately coordinated, calculated and implemented for severe weather avoidance procedures support.”
7/6/99	SFO	The ATCSCC log states: “The program was well planned and conceived based on varying forecasts! ZOA/Bay had no negative comments reference to GDP.”
7/9/99	ORD	The GDP critique states: “ORD went into departure delays at 2142 due to limited departure routes. This caused a backlog of departures on the airport...The ground delay program was issued at 2352 to limit arrivals until departure routes are available and congestion on the airport can be reduced...The program was properly planned and coordinated. The planned program goal of reducing arrival demand in order to avoid gridlock was accomplished. No adverse comments were received from the users.”
7/17/99	SFO	The ATCSCC GDP critique states: “Landing numbers were consistent with the program goal.”
7/18/99	EWR	The departure delays were due to thunderstorms blocking departure routes.
7/19/99	EWR	The ATCSCC GDP critique states: “The program worked to achieve a manageable flow for arrivals and departures and kept delays to a manageable level.”
7/19/99	LGA	The ATCSCC GDP critique states: “The program worked to achieve a manageable flow for arrivals and departures and kept delays to a manageable level.”
7/21/99	EWR	The ATCSCC log states: “the program worked well to keep demand at the airport manageable and assisted with the numerous weather impacts in the enroute system. Delays outbound were kept within reasonable limits as the metered

		flow also regulated the amount of departures."
7/21/99	LGA	The ATCSCC GDP critique states: "The program worked well to keep demand at the airport manageable and assisted with the numerous weather impacts in the enroute system. Delays outbound were kept within reasonable limits as the metered flow also regulated the amount of departures."
7/21/99	PHL	The ATCSCC log states: "the program worked well to keep demand at the airport manageable and assisted with the numerous weather impacts in the enroute system. Delays outbound were kept within reasonable limits as the metered flow also regulated the amount of departures."
7/21/99	SFO	The ATCSCC log states: "The program delivered the anticipated number of arrivals per hour, in addition, airborne holding was kept to a minimum."
7/22/99	EWR	The ATCSCC GDP critique states: "Program worked to meter demand inbound as thunderstorms impacted arrival routes. GDP also allowed for a reduced departure demand that otherwise could have resulted in a gridlock situation. Ground stops were utilized as weather impacted the field. The program was revised following the ground stops and appeared to be effective."
7/22/99	PHL	The ATCSCC GDP critique states: "Program worked to meter demand inbound as thunderstorms impacted arrival routes. GDP also allowed for a reduced departure demand that otherwise could have resulted in a gridlock situation. The program was revised and extended and appeared to be affective."
7/22/99	SEA	The GDP digest states at 1515: "After modeling 15-20 SEA GDPs, I decided on one that capped the delays at 45; the average was 30. Discussed this with ZSE and they were very supportive...The FADT indicates that the delays are more equitably distributed."
7/25/99	SFO	The ATCSCC GDP critique states: "The program delivered arrival numbers from the first two hours consistent with the objective."
7/27/99	ORD	The ATCSCC GDP Critique states: "GDP provided the controls necessary for the ZAU staffing problem, we were able to deliver the arrival as planned, without 'traffic saturation' for ZAU."
7/27/99	SFO	The ATCSCC GDP critique states "The program delivered arrival numbers consistent with the objective."
7/28/99	EWR	The ATCSCC critique at 0057states: "Program achieved the desired results, by controlling arrival numbers and supporting swap."
7/29/99	EWR	The ATCSCC GDP critique states: "Program achieved what was desired from it, departure delays were kept to a minimum

		even with several departure routes being closed for periods."
7/29/99	LGA	The ATCSCC GDP critique states: "Program achieved what was desired from it, departure delays were kept to a minimum even with several departure routes being closed for periods."
8/8/99	SFO	The ATCSCC GDP critique states: "GDP ran well, ...and hourly counts reflect flights were delivered well within the GDP parameters. No arrival nor departure delays reported and CARP were authorized."
8/9/99	SFO	The ATCSCC GDP critique states: "Program ran well, delivering the appropriate numbers of aircraft. ZOA did some minor airborne holding, but it was short-lived."
8/10/99	SFO	The ATCSCC GDP critique states: "GDP ran well, delivering aircraft within program parameters...coordination was timely and complete."

In addition, an avoided GDP event was witnessed at the command center. The specialist running the GDP had the following to say.

Date	**Airport**	**Comment**
3/8/99	ORD	ORD requested a GDP at approximately 1500Z starting at 2200Z and later. They anticipated a drop in the AAR to 60 - 68 due to snow. The ATCSCC advised the major users, UAL and AAL, who started making cancellations. By 1630, the demand had dropped enough to avoid a GDP. The demand drop by hour is as follows:

	1444	1632	Difference
2200	72	54	-18
2300	83	61	-22
0000	75	43	-32
0100	110	80	-30
0200	42	17	-25
0300	70	28	-42
Totals	452	283	-169

This is a drop in the demand by 37 percent over a 6-hour period.

APPENDIX D. Flying Time Analysis and Results

Choosing an Appropriate Metric

Flying Time is defined as a measure of holding or delay absorption associated with arrival runway usage. Using two circles with specified distances from the arrival airport, the traversal time from the outer circle to the inner circle is assessed. From examination of operational data and discussions with knowledgeable personnel, rings at distances of 100 and 40 nmi are appropriate. That traversal time is assessed against two baselines: an idealized, modeled traversal time (based on the flight trajectories estimated by the Collaborative Routing and Coordination Tool (see [Carlson and Rhodes, 1998]), and a Visual Meteorological Condition (VMC) baseline. The VMC baseline is used to evaluate the holding or delay absorption which is the standard air traffic management practice in a certain locality.

The excess arrival flying time (hereafter called simply “excess flying time”) as defined above can be interpreted as follows. The number of flights subjected to excess flying time, and the amount of the excess might increase under a GDP because of uncertainty in the true airport landing capacity, and the reduced controllability of take-off times at origin airports. To the degree that the GDP excess flying time is like the VMC value, then the GDP is successful, since it keeps pressure on the runway via appropriate airborne reservoir. Conceptually, we want to measure whether holding increases as the ATM system seeks to fully utilize scarce runway capacity.

Two sources of data were used: Enhanced Traffic Management System (ETMS) data (for flight time and position information), and Air Traffic Control System Command Center (ATCSCC) program logs (to determine the start time, duration, etc. of the GDP).

This metric is appropriate for two uses: either to (1) assess the performance of GDP-E, relative to GDP, prior to enhancements or, (2) ongoing monitoring of GDP-E. This section generally examines use (1). Regarding terminology in this section, “GDP” refers in the generic, to a ground delay program, the traffic flow management initiative used by the FAA to limit the number of arrivals during reduced capacity situations. The terms “GDP-E” and “post-GDP-E” are used synonymously, meaning the “Enhanced” Ground Delay Program (i.e., including collaboration and data-sharing). The term pre-GDP-E refers to the simpler method, commonly called “Grover Jack.”

A reasonable question is which direction might one expect the metric to go—would we expect more or less excess flying time under pre- vs. post-GDP-E? Musing on causes, it is difficult to say, since some seem in the direction of more, and some of less, excess flying time. GDP-E does a good job of filling unused slots, increasing the pressure on the runway, and perhaps leading to more holding. On the other hand, GDP-E has improved information exchange, plus detailed monitoring of a situation by both the FAA and the airlines. This might lead to less holding. Results of measurements and statistical texts are presented in subsequent sections.

Other Factors and Comparability of Conditions

Many factors influence this metric: weather conditions, air traffic demand level (and direction of flow), operational personnel, and aircraft type mix. These are handled as follows. For aircraft type mix, we have removed lower-performance aircraft, e.g. Beech Aircraft Baron-58, Aerospatiale ATR-72, from consideration. These have a higher variability in speed, and can fly more slowly without actually being in a delay situation. The metric should be considered with this caveat, that not all flights are included in the metric.

Regarding direction of arrival flow, initial tests for various airport arrivals indicate that arrival flows along the differing Standard Terminal Arrival Routes (STARs) exhibit very similar range of traversal time. Other influences: weather, demand level, and time-of-day, are considered and controlled when comparing GDP-E to pre-GDP-E, using paired observations. An important factor, personnel, (i.e., the influence of a style of flow management exhibited by FAA staff) could not be controlled in this study.

Validating the Metric

Figure D-1 presents the distribution of all flights by minutes of flying time during an SFO ground delay program on February 16-17, 1999 (GMT). A program was run beginning at 1700Z on 2/16 until 0645Z on 2/17. A cluster of flights with excess flying time at the beginning of the program can be identified. This is standard practice by flow management, to front-load a program and make sure that no arrival slots go unused in the first hours of a reduced-capacity scenario. The figure provides horizontal lines at 10 and 14 minutes that demarcate *Idealized Flying Time*. Furthermore, note the excess flying time on the left of the chart, when no program was run, reflecting local practice during non-program conditions. Excess flying time during a program can be reduced to a single scalar value: the proportion of large commercial jets with flying time >20 minutes for traversing the 100-to-40 nmi rings. That is, a minimum of +6 minutes above the 14 minute maximum Idealized Flying Time is required before inclusion in the set of delayed flights. This rule is constructed from the plot.

It is expected that if pre- and post-GDP-E are different with respect to holding for the arrival airport, that the metrics would respond accordingly. We would see the difference as a change in the proportion of large commercial jets with excess flying time, and in average and total excess flight minutes.

One important concern is whether ETMS data is an appropriate data source, with respect to demand count. An analysis of recent data for Fort Worth Center and Dallas-Fort Worth tower compared ETMS data to manually tabulated tower counts [Cherniavsky, et al., 1999]. It was determined that ETMS counts are generally within 1 or 2 percent of the tower count.

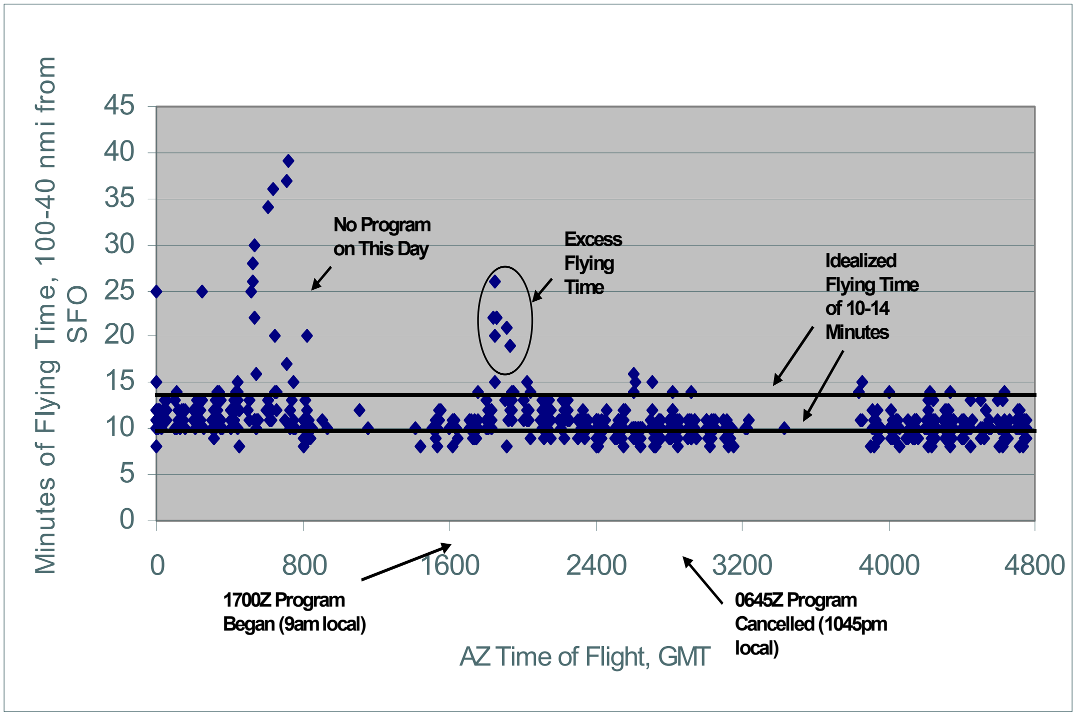

Figure D-1. Flying Time Metric (February 16 - 17, 1999)

Analysis and Results

Matching Days

For an equitable assessment of GDP-E impacts on airborne holding, it is necessary to select pairs of programs with like attributes. One member of the pair should be pre- and one should be post-GDP-E. With such paired observations, several statistical tests are available to assess significant differences.

It is quite a challenge to find comparable situations, and generate enough pairs for significance. In a sense, no two days are exactly alike in the operation of the NAS. However, we have decided on a set of attributes which are probably reasonable for declaring days similar for the purpose at hand. We define the "evaluation interval", as that portion of a GDP that we'd examine – the excess flying time of the arrivals with landing times during this interval. We require, for comparability, that the evaluation interval begin when the GDP begins, and end on or before a program termination or extension. By this requirement we hope for some stability in program dynamics (e.g., weather changes) and the FAA and airline response to changing conditions of weather, landing capacity, and demand.

We require agreement on the following attributes:

1. Same arrival airport
2. Same month of program (+/- 1), might yield similar weather conditions and seasonal schedule

3. Same hour of program start (+/- 1), might yield similar daily schedule
4. Same airport acceptance rate (AAR) (+/- approximately 10%)

Examining Results

Archived ETMS data and ATCSCC logs were scanned and examined to find similar pre- and post-GDP-E samples for comparison[15]. Nine pairs were found.

In Table D-1, the matched pairs are organized across rows, with major column headings of Pre-GDP-E and GDP-E. Within the major column groupings, individual columns contain the following information. (Note that some values are missing, since a few ATCSCC logs were unavailable.)

Some discussion of these Table D-1 entries is appropriate. Excess flying time is likely a function of demand/capacity imbalance. Note that both these quantities are unpredictable and highly variable. Future capacity is very difficult to predict with much certainty. Moreover, characterizing a landing rate as a single value is not a full representation of an hour's airport activity, which often has "peaks and valleys" within an hour. Generally, the FAA consults meteorologists, declares an airport to be IFR for some hours in the future, and associates a known IFR landing rate (based on Engineering Performance Standards and on past experience). It may be that the ATCSCC assessment differs somewhat from the tower supervisor's estimate of landing rate, and hence some variation exists between the FY and the DCC values. Another capacity-related value is the PG, the program goal, the actual value used in the GDP algorithm execution. A consideration of a number of factors is undertaken to decide on this value including weather predictability, weather transition likelihood, predictability of demand, customer (airline) preferences, and feedback. Selection of a PG value may be considered as much an "art" as a science, especially when we see it at variance from the FY and/or DCC values.

Demand is more controllable than capacity, but it is still no mean feat to choreograph just the right amount of flights to show up for arrival services at the right time. Major uncertainties exist in the take-off time of flights during conditions of delayed departure times.

[15] Voluminous data sets had to be scanned for usability, since it is not unusual for recorded operational data to contain data anomalies or missing data.

Definition of Fields for Table D-1

Field	**Definition**
Site	Subject arrival airport
Date	Day/Month/Year of GDP
Evaluation Interval	Subset of the entire GDP duration evaluated in this analysis
PG/FY/DCC	Various values for landing capacity per hour: PG = program goal, the rate used in the GDP FY = landing capacity per the tower facility DCC = ATCSCC assessment of actual landing capacity
AZ's/Hr	Actual arrivals per hour, via the ETMS AZ message
Tower Landing Count	Actual arrivals per hour, as manually tallied by tower personnel
Count of: excess flying time big jets/ all big jets	One of these metrics we seek. As an example, for the entry "23/173 = 13%": 23 = number of large commercial jets with excess flying time landing during evaluation interval 173 = total number of large commercial jets landing during evaluation interval 13% = quotient of the last two values.
Average Excess Minutes per Delayed Flight	Continuing with above example, the 23 delayed flights had an average excess flying time of 11 minutes.
Total Excess Flight Minutes	Continuing with example, the 23 flights had a total of 253 flight minutes excess. (Note that rounding to integers for the last 3 items is performed as a last step.)

Can excess flying time be modeled as a simple relationship of actual landings versus the AAR used in the GDP? One might expect the relationship to be inversely related. When landing count is greater than the stated AAR, then perhaps the true AAR was underestimated, and excess flying time would be low. On the contrary, if landing count is less than or equal to the stated AAR, maybe true AAR was about right or overestimated, and excess flying time would be high. We examined the following relationship, for both pre- and post-GDP-E conditions:

> Percent of big jets with excess flying time = f (Maximum Tower Landing Count for Evaluation Interval minus AAR used in GDP).

We found, in a linear regression analysis a lack of correlation: an R^2 value of 0.03.

This result does not give us great confidence that the further comparison of pre- and post-GDP-E will be very rewarding. Indeed, comparing the pairs (13%, 2%), (12%, 21%),… taken from the Table D-1, we performed two statistical tests. The Paired-T test assumes underlying normality of the data, and tests the null hypothesis that the difference of the pairs is zero. Since we cannot be sure of the assumption of normality, we also applied the Wilcoxon Signed-Rank test, which makes no assumptions about the underlying probability distribution of the data. We apply a two-tailed test, since we care about differences in either direction (GDP-E being either worse or better than pre-GDP-E). Below are the results expressed as p-values or "observed significance." (If the observed significance were, say 0.02, and we desired an answer with 95 percent confidence [i.e., alpha=0.05], we'd reject the null hypothesis, and favor the assertion that there is a systematic effect in the data.)

Test	Observed Significance
Paired-T	0.36
Signed-Rank	0.68

In light of these high values, we fail to reject the null hypothesis, even if alpha were set to a value of 0.10. Similar results were found for the other measures, average and total excess flight minutes.

Summary

We suspect that even with our attempt to match days for comparison, the system is replete with extraneous noise, which would obscure a significant difference, if one indeed exists. Just considering the "physics" of the problem, if the number of flights showing up for arrival services exceeds the safe capacity of a facility (approach control or tower), some type of flow control or delay maneuver *must* be undertaken. But given the complexity of the NAS, and the additional dynamics and complications brought on by a GDP, we are faced with messy, difficult-to-interpret data in our analysis. Given the data analyzed here, we cannot conclude that delay maneuvers for airborne flights are either lower or higher under GDP-E.

This analysis focused on a specific performance metric that attempts to measure the operational impact of GDP-E on delay. The hypothesis that GDP-E is better or worse at airborne holding is inconclusive. This result does not imply that GDP-E does not or has not been found to provide benefits to airspace users and service providers. There are many controlled studies and anecdotal evidence that have shown and continue to show the positive operational impacts of GDP-E.

Table D-1a. Results from Pairing Pre- and Post-GDP-E Samples

	Pre-GDP-E								GDP-E							
Site	**Date**	**Evaluation interval (HHMM, GMT)**	**PG/ FY/ DCC**	**AZs/ hour**	**Tower Landing Count**	**Counts of: excess-flying-time big jets/ all big jets**	**Average Excess Minutes per Delayed Flight**	**Total Excess Flight Minutes**	**Date**	**Evaluation interval (HHMM, GMT)**	**PG/ FY/ DCC**	**AZs/ hour**	**Tower Landing Count**	**Counts of: excess-flying-time big jets/ all big jets**	**Average Excess Minutes per Delayed Flight**	**Total Excess Flight Minutes**
ATL	3/18/98	1900-2359	65/63/ 65	58,57, 59,59, 57	61,73,64, 74,70	23/173=13%	11	253	3/3/99	1900-2359	76/82/ 82	63,65, 71,65, 75	_,_,75,78, 79	4/210=02%	7	28
	1/5/98	1900-2459	66/_/_	58,47, 66,59, 59,61	_	26/212=12%	10	260	12/13/98	1900-0059	70/70, 70,68, 68,72, 72/70	53,60, 65,66, 70,69	59,69,73, 73,76,72	49/231=21%	7	348
BOS	3/9/98	1900-2259	32/28/ 28	26,23, 30,26	31,27,32, 38	32/43=74%	15	494	3/15/99	2000-2359	32/32/ 32	27,27, 25,30	33,30,32, 27	41/53=77%	11	451
	3/19/98	2000-0159	38:2, 36:4/ 38/38	37,31, 37,28, 36,34	34,40,37, 37,39,36	51/102=50%	13	663	3/11/99	2100-0259	40/36/ 40	33,38, 31,33, 31,40	34,42,35, 37,33,38	48/109=44%	10	494
EWR	1/5/98	2000-2359	36/_/_	38,31, 31,23	_	5/70=07%	7	43	12/1/98	2000-2359	40/37/ 40	34,35, 38,37	40,47,48, 50	4/46=09%	26	102

Legend: _ indicates missing values

n:m indicates m replications of the value n

Table D-1b. Results from Pairing Pre- and Post-GDP-E Samples (Concluded)

	Pre-GDP-E								GDP-E							
Site	Date	Evaluation interval (HHMM, GMT)	PG/ FY/ DCC	AZs/ hour	Tower Landing Count	Counts of: excess-flying-time big jets/ all big jets	Average Excess Minute per Delayed Flight	Total Excess Flight Minutes	Date	Evaluation interval (HHMM, GMT)	PG/ FY/ DCC	AZs/ hour	Tower Landing Count	Counts of: excess-flying-time big jets/ all big jets	Average Excess Minute per Delayed Flight	Total Excess Flight Minutes
ORD	3/18/98	1300-1859	68/72/66	62,52,58,56,48,53	58,63,59,59,55,56	12/208=06%	10	117	4/15/99	1300-1859	64/64/64	61,63,63,61,65,66	63,66,63,63,68,76	26/215=12%	9	241
PHL	3/3/98	2000-0159	38/36/38	34,47,27,44,31,40	47,47,36,42,36,40	0/118=0%	0	0	3/14/99	1900-0059	36/36/36	23,29,34,34,28,29	29,37,34,30,32,23	1/113=01%	6	6
	4/19/98	2000-2359	40/40/40	37,40,46,28	36,44,44,35	4/80=05%	4	14	3/3/99	2000-2359	36/36/36	33,48,36,29	26,45,37,31	1/68=01%	1	1
STL	1/6/98	1300-1700	33/_/_	27,31,35,29,33	_	66/100=66%	9	618	12/21/98	1300-1759	32/32/32	30,32,29,27,28	30,35,27,32,28	1/101=01%	4	4

Legend: _ indicates missing values
n:m indicates m replications of the value n

APPENDIX E. EDCT Compliance Data

	Early		On-Time		Late	
Date	No. of Flights	Percentage	No. of Flights	Percentage	No. of Flights	Percentage
Jan-96	276	18.41	637	42.49	586	39.09
Feb-96	2,240	35.29	2,988	47.08	1,119	17.63
Mar-96	3,253	29.48	5,734	51.96	2,049	18.57
Apr-96	2,001	33.12	2,770	45.85	1,271	21.04
May-96	1,870	23.09	4,027	49.72	2,203	27.20
Jun-96	1,024	23.66	2,124	49.08	1,180	27.26
Jul-96	1,310	29.26	1,930	43.11	1,237	27.63
Aug-96	1,375	25.90	2,523	47.52	1,411	26.58
Sep-96	1,893	22.41	4,825	57.13	1,728	20.46
Oct-96	2,314	24.18	5,405	56.48	1,851	19.34
Nov-96	1,778	24.28	3,718	50.78	1,826	24.94
Dec-96	3,109	21.17	7,332	49.93	4,245	28.91
Jan-97	3,293	21.90	7,902	52.54	3,844	25.56
Feb-97	2,098	30.76	3,340	48.97	1,383	20.28
Mar-97	1,712	31.18	2,726	49.65	1,052	19.16
Apr-97	1,548	27.96	2,823	50.99	1,165	21.04
May-97	1,048	25.14	2,336	56.05	784	18.81
Jun-97	1,316	32.83	1,991	49.68	701	17.49
Jul-97	552	24.76	1,336	59.94	341	15.30
Aug-97	1,434	26.79	2,790	52.13	1,128	21.08
Sep-97	827	28.51	1,310	45.16	764	26.34
Oct-97	485	31.58	837	54.49	214	13.93
Nov-97	2,651	27.72	4,900	51.23	2,013	21.05
Dec-97	2,900	29.71	4,628	47.41	2,234	22.88
Jan-98	4,467	30.76	6,890	47.44	3,167	21.81
Feb-98	3,781	34.76	5,216	47.95	1,882	17.30
Mar-98	3,669	31.05	5,753	48.68	2,395	20.27
Apr-98	832	24.84	1,867	55.75	650	19.41
May-98	1,774	24.93	4,033	56.68	1,309	18.40
Jun-98	1,209	22.47	2,912	54.12	1,260	23.42
Jul-98	588	19.76	2,017	67.80	370	12.44
Aug-98	1,011	29.69	1,632	47.93	762	22.38
Sep-98	855	14.29	4,460	74.53	669	11.18
Oct-98	912	15.45	3,982	67.45	1,010	17.11
Nov-98	1,180	15.22	5,611	72.35	964	12.43
Dec-98	1,381	14.11	6,747	68.95	1,658	16.94
Jan-99	2,118	19.70	6,518	60.63	2,115	19.67
Feb-99	1,662	19.66	5,482	64.85	1,309	15.49
Mar-99	1,815	17.65	6,571	63.90	1,898	18.46
Apr-99	2,164	19.23	6,980	62.02	2,111	18.76
May-99	2,814	18.51	9,618	63.26	2,772	18.23
Jun-99	1,479	16.93	5,419	62.02	1,840	21.06
Jul-99	1,376	17.16	5,024	62.65	1,619	20.19